On parameter estimation for Markov sequences and applications in health economics

Anastasiia Motrunich

l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

Université du Maine
Le Mans - Laval

Thèse de Doctorat

Anastasiia MOTRUNICH

Mémoire présenté en vue de l'obtention du
grade de Docteur de l'Université du Maine
sous le label de L'Université Nantes Angers Le Mans

École doctorale : *Sciences et Technologies de l'Information, Mathématiques*

Discipline : *Mathématiques et leurs interactions*
Spécialité : *Statistiques*
Unité de recherche : *Laboratoire Manceau de Mathématiques*

Soutenue le *28 Septembre 2015*

Estimation des paramètres pour les séquences de Markov avec application dans des problèmes médico-économiques

JURY

Rapporteurs :	**Denis BOSQ**, Professeur, Université Pierre et Marie Curie - Paris VI **Gérard DURU,** Directeur de Recherche, CNRS **Mikhaïl NIKULIN**, Professeur, Université Victor Segalen Bordeaux 2
Examinateurs :	**Saïd HAMADENE,** Professeur, Université du Maine **Marina KLEPTSYNA**, Professeur, Université du Maine **Robert LAUNOIS**, Professeur, Réseau d'Evaluation en Economie de la Santé
Directeur de Thèse :	**Youri KOUTOYANTS**, Professeur, Université du Maine
Co-directeur de Thèse :	**Franck MAUNOURY**, Docteur, Statésia

l'université nantes angers le mans
PÔLE DE RECHERCHE ET D'ENSEIGNEMENT SUPÉRIEUR

Université du Maine
Le Mans - Laval

Thèse de Doctorat

Anastasiia MOTRUNICH

Estimation des paramètres pour les séquences de Markov avec application dans des problèmes médico-économiques

On parameter estimation for Markov sequences and applications in health economics

Résumé

Dans la première partie de cette thèse, nous considérons plusieurs problèmes d'estimation de paramètre de dimension finie pour les séquences de Markov dans l'asymptotique des grands échantillons. Le comportement asymptotique des estimateurs bayésiens et les estimateurs obtenus par la méthode des moments sont décrits. Nous montrons que sous les conditions de régularité ces estimateurs sont consistants et asymptotiquement normaux et que l'estimateur bayésien est asymptotiquement efficace. Les estimateur-processus du maximum de vraisemblance un-pas et deux-pas sont étudiés. Ces estimateurs nous permettent de construire des estimateurs asymptotiquement efficaces sur la base de certains estimateurs préliminaires, par exemple, les estimateurs obtenus par la méthode des moments ou l'estimateur de Bayes et la structure de l'estimateur du maximum de vraisemblance un-pas. Nous proposons notamment des processus autorégressifs non linéaires comme exemple et nous illustrons les propriétés de ces estimateurs à l'aide de simulations numériques. Dans la deuxième partie, nous donnons les applications de processus de Markov en économie de la santé. Nous comparons les modèles de Markov homogènes et non-homogènes pour l'analyse coût-efficacité de l'utilisation de pansements transparents contenant un gel de gluconate de chlorhexidine par rapport aux pansements transparents standard. Le pansement antimicrobien protège les accès vasculaire centrale et réduit le risque de bactériémies liées aux cathéters. L'impact de l'approche de modélisation sur la décision d'adopter des pansements antimicrobiens pour les patients gravement malades est discuté.

Mots clés
Séquences de Markov, Estimation de paramètre, Estimation-processus du maximum de vraisemblance un-pas, Estimateur de Bayes, Economie de la santé, Analyse coût-efficacité, Bactériémies liées aux cathéters

Abstract

In the first part of this dissertation we consider several problems of finite-dimensional parameter estimation for Markov sequences in the asymptotics of large samples. The asymptotic behavior of the Bayesian estimators and the estimators of the method of moments are described. It is shown that under regularity conditions these estimators are consistent and asymptotically normal. We show that the Bayesian estimator is asymptotically efficient. The one-step and two-step maximum likelihood estimator-processes are studied. These estimators allow us to construct the asymptotically efficient estimators based on some preliminary estimators, say, the estimators of the method of moments or Bayes estimator and the one-step maximum likelihood estimator structure. We propose particular non-linear autoregressive processes as examples and we illustrate the properties of these estimators with the help of numerical simulations. In the second part we give the applications of Markov processes in health economics. We compare homogeneous and non-homogeneous Markov models for cost-effectiveness analysis of routine use of transparent dressings containing a chlorhexidine gluconate gel pad versus standard transparent dressings. The antimicrobial dressing protects central vascular accesses reducing the risk of catheter-related bloodstream infections. The impact of the modeling approach on the decision of adopting antimicrobial dressings for critically-ill patients is discussed.

Key Words
Markov sequences, Parameter estimation, One-step maximum likelihood estimator-process, Bayesian estimator, Health economics, Cost-effectiveness analysis, Catheter-related bloodstream infection

Acknowledgement

First and foremost, I wish to express my sincere gratitude to my advisor, Prof. Yury A. Kutoyants, for his great contribution to this research, for clarity of explications and continuous support in the difficult moments of this writing. For his presence and willingness to discuss and help on Saturdays. His immense knowledge and patience guided me through the maze.

I would like to acknowledge, with deepest gratitude, my co-advisor , Dr. Franck Maunoury, for his excellent guidance and contribution to the health economical part of this dissertation.

Especial thanks to my colleague from Statésia Carole Maunoury for an excellent atmosphere giving me the possibility and time to work on this subject.

I would also like to thank the members of my dissertation committee : Prof. Denis Bosq, Prof. Mikhaïl Nikulin, Prof. Gérard Duru, Prof. Robert Launois, Prof. Marina Kleptsyna, Prof. Saïd Hamadène for their assistance and great interest in this topic.

I am also grateful to all the members of "Probability and Statistics" laboratory and the Department of Mathematics at the University of Maine for providing me with all the necessary facilities for the research.

I would like to acknowledge Dr. Maria Palka-Santini for her expert medical explications concerning the application part of my research.

Finally, I appreciate the financial support from Statésia. Also this work was done under partial financial support of the grant of RSF number 14-49-00079.

And last, but not least, I would like to thank my family for encouraging me over these three years. And my fiancé, Dmitry Tolstoy, thank you, for your love.

Contents

Introduction

Cette thèse est consacrée aux problèmes de l'estimation des paramètres de séquences de Markov et l'utilisation des modèles de Markov dans le domaine de l'économie de la santé. Elle est donc composée de deux parties.

La première est dédiée aux problèmes de l'estimation des paramètres multidimensionnels de séquences de Markov $X^n = (X_0, X_1, \ldots, X_n)$ présentées sous la forme de processus autorégressif non linéaire d'ordre 1. Pour ce modèle, nous décrivons le comportement asymptotique de plusieurs types d'estimateurs. Nous montrons la consistance et la normalité asymptotique des estimateurs bayésiens et des estimateurs obtenus par la méthode des moments. Le principal résultat de la première partie est la construction des estimateur-processus du maximum de vraisemblance un-pas et deux-pas (one and two-step MLE-processes). Les résultats sont illustrés par des simulations numériques, qui confirment l'applicabilité de ces méthodes.

La deuxième partie est dédiée aux problèmes d'application des modèles de Markov dans le domaine médico-économique. Le but de cette partie est d'évaluer le rapport coût-efficacité de l'utilisation de pansements transparents contenant un gel de gluconate de chlorhexidine chez les patients gravement malades en se concentrant sur la perspective de l'unité de soins intensifs. Pour mesurer l'impact sur les résultats d'analyse coût-efficacité, deux approches de modélisation ont été considérées : le modèle de Markov homogène et le modèle de Markov non-homogène. Ces deux approches différentes sont fondées sur les chaînes de Markov d'ordre 1, qui est une suite de variables aléatoires $X^n = (X_0, X_1, \ldots, X_n)$. Elle prend ses valeurs dans l'espace dénombrable d'états prédéfinis et les probabilités de transition entre ces états dépendent de la matrice de transition.

Nous abordons ensuite en détail le contenu de ces deux parties.

Dans le premier chapitre nous considérons plusieurs problèmes de l'estimation des paramètres de dimension finie pour les séquences de Markov dans l'asymp-

totique de grands échantillons. Supposons que $\pi(\vartheta, X_{j-1}, X_j)$ est une densité de transition qui dépend d'un paramètre inconnu $\vartheta \in \Theta \subset R^d$.

Le modèle considéré est une série temporelle non linéaire $(X_j)_{j\geq 0}$ qui est décrit par l'équation

$$X_j = S(\vartheta, X_{j-1}) + \varepsilon_j, \quad j = 1, 2, \ldots,$$

où les variables aléatoires $(\varepsilon_j)_{j\geq 1}$ sont i.i.d. avec une fonction de densité régulière $g(x)$. La fonction $S(\vartheta, x)$ est supposée être connue et régulière par rapport à $\vartheta \in \Theta \subset R^d$, où Θ est un ensemble ouvert, borné et convexe.

Le processus $(X_j)_{j\geq 0}$ a la densité de transition

$$\pi(\vartheta, x, x') = g(x' - S(\vartheta, x)).$$

Nous sommes intéressés par les deux méthodes d'estimation des paramètres. Le comportement asymptotique des estimateurs bayésiens $\tilde{\vartheta}_n$ et les estimateurs obtenus par la méthode des moments $\bar{\vartheta}_n$ sont décrits. Sous les conditions de régularité nous montrons que ces estimateurs sont consistants et asymptotiquement normaux

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta\right) \Rightarrow \mathcal{N}\left(0, \mathbb{I}(\vartheta)^{-1}\right),$$

$$\sqrt{n}\left(\bar{\vartheta}_n - \vartheta\right) \Rightarrow \mathcal{N}\left(0, \mathbb{B}(\vartheta)\right),$$

où $\mathbb{I}(\vartheta)$ est une matrice d'information de Fisher et $\mathbb{B}(\vartheta)$ est une matrice de covariance.

L'étude des estimateurs bayésiens dans un certain sens est une suite de travail de Varakin et Veretennikov (2002). Ils ont décrit les propriétés asymptotiques de l'estimateur du maximum de vraisemblance dans le cas unidimensionnel ($d = 1$) à l'aide du résultat général d'Ibragimov et Hasminskii (1981).

Nous avons montré que dans le cas régulier l'estimateur bayésien possède les mêmes propriétés asymptotiques que l'estimateur du maximum de vraisemblance.

Ensuite nous étudions l'estimateur obtenu par la méthode des moments. Dans le cas régulier il est consistant et asymptotiquement normal, mais avec la covariance limite non optimale.

Les résultats obtenus dans ce chapitre sont illustrés par plusieurs exemples numériques. Nous avons étudié l'exemple de Varakin et Veretennikov (2002), où l'estimateur du maximum de vraisemblance peut être représenté sous la forme explicite. Ensuite, nous avons étudié les deux autres modèles non-linéaires qui vont servir pour la suite dans la construction des estimateurs du maximum de vraisemblance un-pas et deux-pas.

Dans le deuxième chapitre, les estimateurs du maximum de vraisemblance un-pas et deux-pas sont étudiés. Ces estimateurs sont récemment introduits par Y. Kutoyants pour les processus de diffusion ergodiques. Les estimateurs du maximum de vraisemblance un-pas et deux-pas nous permettent de construire des estimateurs asymptotiquement efficaces sur la base de certains estimateurs préliminaires, par exemple, les estimateurs obtenus par la méthode des moments ou l'estimateur de Bayes et la structure de l'estimateur du maximum de vraisemblance un-pas.

L'estimation du maximum de vraisemblance un-pas se fait en deux étapes. Premièrement, nous fixons une période d'apprentissage $X^N = (X_0, X_1, \ldots, X_N)$ des observations $X^n = (X_0, \ldots, X_n)$, où $N = \left[n^\delta\right]$ (N est une partie entière de n^δ) et nous estimons le paramètre inconnu par les observations dans l'intervalle d'apprentissage. Le choix du paramètre $\delta < 1$ va être discuté. Cette période d'apprentissage correspond à une partie d'observations relativement courte.

Ensuite nous utilisons l'idée d'estimateur du maximum de vraisemblance un-pas pour construire un estimateur-processus $\vartheta_n^\star = \left(\vartheta_{k,n}^\star, k = N+1, \ldots, n\right)$ qui est asymptotiquement équivalent à l'estimateur du maximum de vraisemblance.

Introduisons en premier lieu la variable $s \in (\tau_\delta, 1]$, où $\tau_\delta = n^{-1+\delta} \to 0$ et $k = [sn]$. Nous pouvons écrire $\vartheta_{k,n}^\star = \vartheta_{s,n}^\star$ et considérer l'estimateur-processus $\vartheta_n^\star = \left(\vartheta_{s,n}^\star, s \in (\tau_\delta, 1]\right)$. Notre but est ainsi de construire l'estimateur-processus $\vartheta_n^\star$ qui est asymptotiquement optimal pour toutes $s \in (\tau_\delta, 1]$.

L'estimateur du maximum de vraisemblance un-pas peut être calculé avec la procédure suivante :

$$\vartheta_{s,n}^* = \bar{\vartheta}_N + \frac{1}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \Delta_k(\bar{\vartheta}_N, X_N^k),$$

où

$$\Delta_k(\vartheta, X_N^k) = \frac{1}{\sqrt{k}} \sum_{j=N+1}^{k} \dot{\ell}(\vartheta, X_{j-1}, X_j),$$

est une fonction-score et $k = [sn] \to \infty$.

Sous les conditions de régularité nous montrons que

$$\sqrt{k}(\vartheta_{s,n}^* - \vartheta) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta\right)^{-1}\right).$$

Donc l'estimateur-processus $\vartheta_n^\star = \left(\vartheta_{s,n}^\star, \tau_\delta < s \leq 1\right)$ pour toutes $s \in (\tau_\delta, 1]$ est asymptotiquement normal avec la matrice de covariance asymptotiquement optimale.

Pour avoir une période d'apprentissage plus courte, nous introduisons un estimateur intermédiaire en plus et l'estimateur du maximum de vraisemblance deux-pas $\vartheta^{\star\star}_{s,n} = \left(\vartheta^{\star\star}_{k,n}, k = N+1, \ldots, n\right)$. Cette procédure nous conduit à l'estimateur-processus qui est aussi asymptotiquement efficace.

L'estimation du maximum de vraisemblance deux-pas se fait en trois étapes. Premièrement, à la base des observations $X^N = (X_0, X_1, \ldots, X_N)$ nous obtenons l'estimateur préliminaire $\vartheta_{N,1}$ qui est asymptotiquement normal. Ensuite, nous introduisons le second estimateur préliminaire

$$\bar{\vartheta}_{k,2} = \bar{\vartheta}_{N,1} + \frac{1}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_{N,1}\right)^{-1} \Delta_k(\bar{\vartheta}_{N,1}, X^k),$$

où $\Delta_k(\vartheta, X^k)$ est la fonction-score définie ci-dessus.

L'estimateur-processus du maximum de vraisemblance deux-pas $\vartheta^{\star\star}_{s,n} = \left(\vartheta^{\star\star}_{k,n}\right)$, où $k = N+1, \ldots, n$ nous construisons en utilisant ce deuxième estimateur préliminaire :

$$\vartheta^{\star\star}_{s,n} = \bar{\vartheta}_{k,2} + \frac{1}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_{k,2}\right)^{-1} \Delta_k(\bar{\vartheta}_{k,2}, X^k).$$

Sous les conditions de régularité nous montrons que l'estimateur-processus $\vartheta^{\star\star}_{s,n}$ est asymptotiquement normal

$$\sqrt{k}(\vartheta^{\star\star}_{s,n} - \vartheta) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta\right)^{-1}\right)$$

et sa matrice de covariance est asymptotiquement optimale.

L'avantage principal des estimateur-processus du maximum de vraisemblance un-pas et deux-pas est qu'ils sont asymptotiquement efficaces pour chaque s et en même temps facile à calculer. Notons que l'estimateur-processus $\vartheta^{\star}_{k,n}$ peut être écrit dans la forme récurrente :

$$\vartheta^{\star}_{k+1,n} = \frac{k}{k+1}\vartheta^{\star}_{k,n} + \frac{1}{k+1}\bar{\vartheta}_N + \frac{1}{k+1}\mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \dot{\ell}\left(\bar{\vartheta}_N, X_k, X_{k+1}\right).$$

Pour illustrer les résultats de ce chapitre, nous proposons des processus autorégressifs non-linéaires comme les modèles d'observations. Les estimateur-processus du maximum de vraisemblance un-pas et deux-pas sont obtenus dans les deux exemples avec les simulations numériques. Dans le premier cas nous considérons un estimateur du maximum de vraisemblance comme l'estimateur préliminaire et nous cherchons à l'améliorer à l'aide de la procédure du maximum de vraisemblance un-pas. Dans le deuxième cas nous proposons l'estimateur obtenu par la méthode

des moments comme l'estimateur préliminaire et nous cherchons à l'améliorer à l'aide de la procédure du maximum de vraisemblance un-pas et deux-pas.

Dans le troisième chapitre, nous considérons les applications de processus de Markov en économie de la santé. L'intérêt médical de ce travail est la gestion des patients gravement malades qui ont eu l'insertion des cathéters intravasculaires (cathéters veineux centraux et cathéters artériels) dans l'unité de soins intensifs. L'insertion des cathéters intravasculaires peut conduire à des complications infectieuses graves, y compris les bactériémies liées aux cathéters, car ils sont une porte d'entrée pertinente pour les micro-organismes dans la circulation sanguine. C'est une complication relativement fréquente (1-5 épisodes / 1000 jours-cathéter) et potentiellement mortelle observée chez les patients gravement malades dans les unités de soins intensifs.

Nous comparons les modèles de Markov homogène et non-homogène de l'analyse coût-efficacité de l'utilisation de pansements transparents contenant un gel de gluconate de chlorhexidine par rapport aux pansements transparents standard. Le pansement antimicrobien protège les accès vasculaire centraux et réduit le risque de bactériémies liées aux cathéters. Les deux modèles simulent les différentes trajectoires observables de la santé des patients relative au risque de contracter une infection liée aux cathéters et évaluent l'incertitude autour des estimations des essais cliniques.

Dans le modèle de Markov homogène nous considérons l'ensemble de 6 états de santé de patients $\mathcal{E} = (E_1, E_2, \ldots, E_6)$:

$$
\begin{aligned}
E_1 &= \{\text{Aucune infection liée aux cathéters}\}\,; \\
E_2 &= \{\text{Infection liée aux cathéters}\}; \\
E_3 &= \{\text{Dermatite de contact}\}; \\
E_4 &= \{\text{Changement du pansement}\}; \\
E_5 &= \{\text{Décharge}\}; \\
E_6 &= \{\text{Décès}\}\ .
\end{aligned}
$$

Dans le modèle de Markov non-homogène nous considérons 8 états de santé de

patients $\mathcal{E} = (E_1, E_2, \ldots, E_8)$:

$E_1 = \{$Insertion d'un premier cathéter, aucune infection liée aux cathéters$\}$;
$E_2 = \{$Insertion d'un nouvel cathéter, aucune infection liée aux cathéters$\}$;
$E_3 = \{$Infection liée aux cathéters sans l'insertion d'un nouvel cathéter$\}$;
$E_4 = \{$Infection liée aux cathéters avec l'insertion d'un nouvel cathéter$\}$;
$E_5 = \{$Dermatite de contact$\}$;
$E_6 = \{$Changement du pansement$\}$;
$E_7 = \{$Décharge$\}$;
$E_8 = \{$Décès$\}$.

Les états Décharge et Décès sont les états absorbants. Nous considérons l'ensemble de probabilités $\Theta = (\pi(E_l, E_m))_{6\times 6}$ ou $\Theta = (\pi(E_l, E_m))_{8\times 8}$ comme les matrices de transition inconnues. Les estimateurs de ces matrices $\hat{\pi}(E_l, E_m)$ sont été obtenus à l'aide de base de données en vie réelle. La matrice de probabilités de transition peut être considérée comme le paramètre multidimensionnel provenant de cette base de données. Et tous les calculs sont fait à partir de ces estimateurs de probabilités de transition $\hat{\pi}(E_l, E_m)$.

Ces nouveaux modèles sont profondément différents de toutes les évaluations économiques antérieures de pansements antimicrobiens concernant la prévention des infections liées aux cathéters. Ces évaluations ont utilisé des modèles d'arbre décisionnel représentant les choix thérapeutiques.

L'impact de l'approche de modélisation sur la décision d'adopter des pansements antimicrobiens pour les patients gravement malades est discuté.

Les résultats de ces chapitres ont fait l'objet de publications et présentations orales.

Les articles :

1. Motrunich, A. On parameter estimation for Markov sequences. Soumis, 2015.
2. Kutoyants, Y.A. and Motrunich, A. On multi-step MLE-process for Markov sequences. Soumis à Metrika. (Accepté pour publication avec les modifications légères) 2015.
3. Maunoury, F., Motrunich, A., Palka-Santini, M., Bernatchez, S.F., Ruckly, S., Timsit, J.F. Cost-effectiveness analysis of a transparent antimicrobial dressing for managing central venous and arterial catheters in intensive care unit. PLoS ONE 2015 ; 10(6) : e0130439. doi :10.1371/journal.pone.0130439.

Les posters et présentations orales :

1. On parameter estimation for Markov sequences. *Asymptotical Statistics of Stochastic Processes. Workshop : S.A.P.S. X*, Le Mans, 2015.

2. Modeling cost-effectiveness of antimicrobial dressing for preventing catheter-related bloodstream infection : homogeneous vs non- homogeneous Markov approaches. *International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 17th Annual European Congress*, Amsterdam, 2014.

3. Cost-effectiveness analysis of an antimicrobial transparent dressing for protecting central vascular accesses in critically ill patients versus standard transparent dressing in France : A comparison of two modeling approaches : Decision-tree versus non-homogeneous Markov model. *International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 17th Annual European Congress*, Amsterdam, 2014.

4. Non-homogeneous cost-effectiveness modeling of a new CHG-dressing for preventing catheter-related bloodstream infections for patients in intensive care units. *International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 16th Annual European Congress*, Dublin, 2013.

5. Cost-effectiveness of the TLC-NOSF dressing in venous leg ulcers. *International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 15th Annual European Congress*, Berlin, 2012.

Chapter 1

On Parameter Estimation for Markov Sequences

1.1 Introduction

We consider the problem of parameter estimation in the case of observations of Markov sequence $X^n = (X_0, X_1, X_2, \ldots, X_n)$. We suppose that the transitional density $\pi(\vartheta, X_{j-1}, X_j)$ depends on some unknown finite-dimensional parameter θ and we study the properties of the estimators of this parameter in the asymptotic of large samples ($n \to \infty$). It is known that under regularity conditions the maximum likelihood estimator (MLE) $\hat{\vartheta}_n$ is consistent, asymptotically normal

$$\sqrt{N}\left(\hat{\vartheta}_N - \vartheta\right) \Rightarrow \mathcal{N}\left(0, \mathbb{I}(\vartheta)^{-1}\right)$$

and asymptotically efficient (see, [13], [17]). Here $\mathbb{I}(\vartheta)$ is Fisher information matrix.

In our work we are interested by two methods of the parameter estimation. The first one is Bayesian and the second is the method of moments. In both cases we show that the corresponding estimators are consistent and asymptotically normal.

The model of observations is a nonlinear time series $(X_j)_{j\geq 0}$ satisfying the relation

$$X_j = S(\vartheta, X_{j-1}) + \varepsilon_j, \quad j = 1, 2, \ldots \tag{1.1}$$

and the initial value X_0 is given too. The random variables $(\varepsilon_j)_{j\geq 1}$ are i.i.d. with smooth density function $g(\cdot)$. The process $(X_j)_{j\geq 0}$ has a transition density

$$\pi(\vartheta, x, x') = g(x' - S(\vartheta, x)).$$

It depends on the parameter ϑ and defines the probability of reaching the state x' after sojourning in the state x. The parameter ϑ takes its values in some open bounded set $\Theta \subset R^d$.

We consider two estimators. The first one is constructed following Bayesian approach, i.e., we suppose that the unknown parameter is a random variable with known *prior* density $p(\vartheta), \vartheta \in \Theta$. For the simplicity of exposition we take the quadratic loss function $W(u) = |u|^2$ and therefore the Bayesian estimator (BE) $\tilde{\vartheta}_n$ is the conditional expectation $\tilde{\vartheta}_n = \mathbf{E}(\vartheta | X^n)$. It has the representation

$$\tilde{\vartheta}_n = \left(\int_\Theta p(\theta) V(\theta, X^n) \, \mathrm{d}\theta \right)^{-1} \int_\Theta \theta p(\theta) V(\theta, X^n) \, \mathrm{d}\theta,$$

where the likelihood function

$$V(\vartheta, X^n) = \pi_0(X_0) \prod_{j=1}^{n} \pi(\theta, X_{j-1}, X_j), \qquad \vartheta \in \Theta. \tag{1.2}$$

Here $\pi_0(x)$ is the density of the initial value X_0.

We suppose in this work that the process $(X_j)_{j \geq 0}$ is geometrically mixing and has invariant distribution with the density $\pi^*(\vartheta, x)$. For simplicity of exposition we put $\pi_0(x) = \pi^*(\vartheta, x)$. In this case the process $(X_j)_{j \geq 0}$ is stationary.

(H) The function $\pi(\vartheta, x, x')$ is two times continuously differentiable in ϑ and the derivative of the function

$$\ell(\vartheta, x, x') = \ln \pi(\vartheta, x, x')$$

has polynomial majorants. Moreover we suppose that the invariant density has polynomially decreasing tails, such that the moments of the corresponding functions used in the proof exist.

Of course, this means that the functions $g(\cdot) \in \mathcal{C}^2$ and $S(\cdot, \cdot) \in \mathcal{C}^2_\vartheta$ with the corresponding conditions on their derivatives.

Our work is a continuation of the work by Varakin and Veretennikov [17], where the asymptotic properties of the MLE of the one-dimensional parameter ϑ are described with the help of the general results by Ibragimov and Hasminskii [3].

We suppose that all conditions of regularity imposed on the function $f(\vartheta, x, x')$ in [17] and providing the ergodicity, the existence of the unique invariant measure $\pi^*(\vartheta, x, x')$ satisfying the mentioned above condition on the tails hold. The conditions of Theorem 1 in [17] are supposed to be fulfilled too. Moreover, our proofs of the asymptotic properties of the likelihood ratio process (Lemmas 1-3) follow

the main steps of the corresponding proofs in [17]. The only difference is in the dimension of the unknown parameter.

We have the *law of large numbers with rate*: for any $p > 0$

$$\mathbf{E}_\vartheta \left| \frac{1}{n} \sum_{j=1}^{n} \left[h\left(\vartheta, X_j\right) - \mathbf{E}_\vartheta h\left(X_j\right) \right] \right|^p \leq \frac{C}{n^{p/2}}. \tag{1.3}$$

These regularity conditions allow us, for example, differentiate w.r.t. ϑ under sign of mathematical expectation. The Fisher information matrix is

$$\mathbb{I}(\vartheta) = \mathbf{E}^*_\vartheta \left[\dot{\ell}\left(\vartheta, X_0, X_1\right) \dot{\ell}\left(\vartheta, X_0, X_1\right)^{\mathbb{T}} \right],$$

where the dot means the derivation w.r.t. ϑ and $\mathbb{T}$ means the transpose of a matrix. The mathematical expectation here is w.r.t. the invariant measure $\pi^*\left(\vartheta, \cdot\right)$.

We suppose that this matrix is uniformly in $\vartheta \in \Theta$ non-degenerate and bounded

$$0 < \inf_{\vartheta \in \Theta} \inf_{|\lambda|=1} \lambda^{\mathbb{T}} \mathbb{I}(\vartheta) \lambda, \tag{1.4}$$

$$\sup_{\vartheta \in \Theta} \sup_{|\lambda|=1} \lambda^{\mathbb{T}} \mathbb{I}(\vartheta) \lambda < \infty. \tag{1.5}$$

Here $\lambda \in R^d$.

The maximum likelihood estimator we introduce as usual by the equation

$$V(\hat{\vartheta}_n, X^n) = \sup_{\vartheta \in \Theta} V(\vartheta, X^n). \tag{1.6}$$

If this equation has more than one solution then we can take any of them as the MLE.

We introduce as well the log-likelihood ratio function

$$L(\vartheta, X^n) = \ln \pi^*(\vartheta, X_0) + \sum_{j=1}^{n} \ln \pi(\vartheta, X_{j-1}, X_j). \tag{1.7}$$

It is known that under the regularity conditions the MLE has the following properties

1. *It is consistent* (here and throughout this work consistent means consistent in probability): $\hat{\vartheta}_n \longrightarrow \vartheta$.

2. *Asymptotically normal*:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \Longrightarrow \mathcal{N}(0, \mathbb{I}(\vartheta)^{-1}), \tag{1.8}$$

where $\mathbb{I}(\vartheta)$ is the Fisher information matrix (defined above).

3. *Asymptotically efficient*, i.e., it satisfies the relation: for all $\vartheta_0 \in \Theta$

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta}\mathbf{E}_\vartheta W\left(\sqrt{n}\left(\hat{\vartheta}_n - \vartheta\right)\right) = \mathbf{E}W\left(\mathbb{I}(\vartheta_0)^{-1/2}\zeta\right). \tag{1.9}$$

The same time for all estimators ϑ_n the following Hajek-Le Cam's type lower bound

$$\underline{\lim}_{\delta\to 0}\underline{\lim}_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta}\mathbf{E}_\vartheta W\left(\sqrt{n}\left(\vartheta_n - \vartheta\right)\right) \geq \mathbf{E}W\left(\mathbb{I}(\vartheta_0)^{-1/2}\zeta\right) \tag{1.10}$$

holds [18].

Here $W(u) = |u|^p, u \in R^d$ with $p > 0$ is a loss function (see, e.g., [3]) and ζ is a Gaussian vector $\zeta \sim \mathcal{N}(0, \mathbb{J})$, $\mathbb{J}$ is a unit $d \times d$ matrix.

The proofs of these properties can be found in [13](properties 1 and 2) and in [17] (property 3) in the one-dimensional case ($d = 1$). See [16] for the discussions of related works. Note that there one can find the study of the properties of another well studied estimator called *conditional least square* estimator.

Our goal is to show that the Bayesian estimator has the same asymptotic properties.

The second estimator studied in this work is the estimator of the method of moments (EMM) $\bar{\vartheta}_n$. Recall that these estimators (under regularity conditions) are consistent and asymptotically normal, but their limit covariance function is different of the inverse Fisher information matrix and therefore these estimators are not asymptotically efficient.

Nevertheless the study of these estimators can be interesting because the construction of them can be more simple than that of the MLE or BE. Moreover, we show in the forthcoming work [9] that using these estimators and multi-step procedure it is possible to obtain estimators asymptotically equivalent to MLE and BE (see [8] where this construction is discussed in details for some models of stochastic processes).

This is the first part of the study devoted to the construction of multi-step MLE-processes for Markov sequences [9]. The studied here two estimators are supposed to be used in the next work as preliminary estimators. The goal is to

propose an estimator-process $\vartheta^{\star}_{j,n}, j = N, \dots, n$, which can be easily calculated for all $N \leq j \leq n$ with $N = n^{\delta}, \delta < 1$ and which has asymptotically optimal properties.

1.2 Motivation

Suppose we have some discrete-time random process (in economics, in medicine, etc.) that we can describe with a well-known autoregressive model (AR) :

$$AR(1): \; X_{j+1} = \vartheta X_j + \varepsilon_{j+1}, \quad j = 1, 2, \dots$$

or

$$AR(p): \; X_{j+1} = \sum_{l=1}^{p} \vartheta_l X_{j-l} + \varepsilon_{j+1}, \quad j = 1, 2, \dots .$$

Here ϑ_l are the parameters of the model and ε_j is white Gaussian noise, i. e., $\mathbf{E}\varepsilon_j = 0, \mathbf{E}\varepsilon_j^2 = \sigma^2, \mathbf{E}\varepsilon_j\varepsilon_i = 0, i \neq j$.

There are many ways to estimate these parameters. Recall some results of the methods of maximum likelihood estimator, Bayes estimator and estimator method of moments.

Maximum Likelihood Estimator

Let us denote $\hat{\vartheta}_n$ the maximum likelihood estimator of the true value ϑ. Recall the definition of this estimator

$$V(\hat{\vartheta}_n, X^n) = \sup_{\vartheta \in \Theta} V(\vartheta, X^n). \tag{1.11}$$

If ε_j are i.i.d. r.v.'s $\mathcal{N}(0, \sigma^2)$ and we have the $AR(1)$ model then the MLE is easy to calculate:

$$\hat{\vartheta}_n = \frac{\sum_{j=1}^{n-1} X_{j+1} X_j}{\sum_{j=1}^{n-1} X_j^2}.$$

The properties of MLE follow from the representation:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) = \frac{\frac{1}{\sqrt{n}} \sum_{j=1}^{n-1} X_j \varepsilon_{j+1}}{\frac{1}{n} \sum_{j=1}^{n-1} X_j^2} \Longrightarrow \mathcal{N}\left(0, 1 - \vartheta^2\right).$$

This limit follows from the law of large numbers

$$\frac{1}{n}\sum_{j=1}^{n-1} X_j^2 \longrightarrow \mathbf{E}_\vartheta X_1^2 = \frac{\sigma^2}{1-\vartheta^2}$$

and from the Central Limit Theorem

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n-1} X_j\varepsilon_{j+1} \Longrightarrow \mathcal{N}\Big(0, \frac{\sigma^4}{1-\vartheta^2}\Big).$$

The Fisher information in this problem is

$$\mathbb{I}\,(\vartheta) = \frac{1}{1-\vartheta^2}$$

and therefore the MLE is asymptotically efficient.

Hajek-Le Cam's lower bound

The asymptotic efficiency is understood in the following way. For all estimators ϑ_n and all $\vartheta_0 \in \Theta$ we have

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta} \mathbf{E}_\vartheta W\left(\sqrt{n}\,(\vartheta_n-\vartheta)\right) \geq \mathbf{E}W\left(\mathbb{I}(\vartheta_0)^{-1/2}\zeta\right). \tag{1.12}$$

This is Hajek-Le Cam's lower bound on the risks of all estimators, which is valid under regularity conditions for any statistical model. Here $W\,(u)$ is some loss function and $\zeta \in \mathcal{N}(0, \mathbb{J})$.

It is known that for the $AR(1)$ model with $|\vartheta| < 1$ the MLE $\hat{\vartheta}_n$ is asymptotically efficient, i.e., it satisfies the relation: for all $\vartheta_0 \in \Theta$

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta} \mathbf{E}_\vartheta W\left(\sqrt{n}\left(\hat{\vartheta}_n-\vartheta\right)\right) = \mathbf{E}W\left(\mathbb{I}(\vartheta_0)^{-1/2}\zeta\right).$$

Bayes Estimator

In our work we will study the BE that can be calculated much easier than the MLE in some cases. Note that in the regular cases the asymptotic properties of the BE and MLE are equivalent [3].

Suppose that the unknown parameter ϑ is a random vector with the density *a priory* $p(\vartheta), \vartheta \in \Theta$.

Then the BE $\tilde{\vartheta}_n$ is calculated as follows

$$\tilde{\vartheta}_n = \frac{\int_\Theta \vartheta p(\vartheta) V(\vartheta, X^n) \, \mathrm{d}\vartheta}{\int_\Theta p(\vartheta) V(\vartheta, X^n) \, \mathrm{d}\vartheta}. \tag{1.13}$$

This estimator under regularity conditions is consistent, asymptotically normal

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}(\vartheta)^{-1}\right)$$

and asymptotically efficient [3].

Suppose that we observe the AR(1) model, where the unknown parameter $\vartheta \in [\alpha, \beta]$ and $-1 < \alpha < \beta < 1$ is a r. v. with the prior density $p(\vartheta), \vartheta \in [\alpha, \beta]$. From (1.13), the BE $\tilde{\vartheta}_n$ has the following representation:

$$\tilde{\vartheta}_n = \frac{\int_\alpha^\beta \vartheta p(\vartheta) \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^n (X_j - \vartheta X_{j-1})^2\right) \mathrm{d}\vartheta}{\int_\alpha^\beta p(\vartheta) \exp\left(-\frac{1}{2\sigma^2}\sum_{j=1}^n (X_j - \vartheta X_{j-1})^2\right) \mathrm{d}\vartheta}.$$

We have

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, 1 - \vartheta^2\right).$$

Estimator of the Method of Moments

Let us consider the same example as previously. Suppose that we have the observations X^n of $AR(1)$. Then by the law of large numbers

$$\frac{1}{n}\sum_{j=1}^n X_j^2 \longrightarrow \frac{\sigma^2}{1 - \vartheta^2}.$$

Hence the estimator of the method of moments is

$$\bar{\vartheta}_n = \sqrt{\left(1 - \frac{\sigma^2}{\frac{1}{n}\sum_{j=1}^n X_j^2}\right)_+} \longrightarrow \vartheta.$$

Here $A_+ = \max(A, 0)$. Usually the limit variance of the EMM is greater than that of the MLE or BE, but in some problems the calculation of this estimator can be easier and its use by this reason can be preferable.

Recall the definition of EMM. Suppose that we have some autoregressive time series

$$X_{j+1} = S(\vartheta, X_j) + \varepsilon_{j+1}, \quad j = 0, 1, 2, \ldots$$

with invariant density function $\pi^*(\vartheta, x)$.

Let us take such function $q(x)$ that the limit function $m(\vartheta)$ defined by the relation

$$\frac{1}{n}\sum_{j=1}^{n} q(X_j) \longrightarrow m(\vartheta) = \int q(x)\,\pi^*(\vartheta, x)\,\mathrm{d}x$$

is strictly monotone, i.e., the equation $m(\vartheta) = t$ has a unique solution $\vartheta = h(t)$. Then we define the EMM

$$\bar{\vartheta}_n = h\left(\frac{1}{n}\sum_{j=1}^{n} q(X_j)\right) \longrightarrow h(m(\vartheta)) = \vartheta.$$

This estimator is asymptotically normal

$$\sqrt{n}\left(\bar{\vartheta}_n - \vartheta\right) \Longrightarrow \mathcal{N}(0, D(\vartheta))$$

with some covariant matrix $D(\vartheta)$.

1.3 Bayes estimator

Suppose that we observe the time series (1.1), where the unknown parameter $\vartheta \in \Theta$ is a random vector with the prior density $p(\vartheta), \vartheta \in \Theta$. The function $p(\cdot)$ is continuous, bounded and positive. We are interested by the behavior of Bayes estimator for the quadratic loss function, which has the following representation:

$$\tilde{\vartheta}_n = \frac{\int_\Theta \vartheta p(\vartheta)\, V(\vartheta, X^n)\,\mathrm{d}\vartheta}{\int_\Theta p(\vartheta)\, V(\vartheta, X^n)\,\mathrm{d}\vartheta}.$$

Recall that as usual in the regular cases the asymptotic properties of the BE are equivalent to the properties of the MLE [3].

Theorem 1 *Let the conditions of regularity be fulfilled. Then the BE $\tilde{\vartheta}_n$ is consistent, asymptotically normal*

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta_0) \Longrightarrow N(0, \mathbb{I}(\vartheta_0)^{-1}) \tag{1.14}$$

and asymptotically efficient for the polynomial loss functions.

Proof. We have to check the conditions H1-H4 of the Theorem 3.2.1 in [3], but first we recall the main steps of the proof. Note that we follow the main steps of the proof given in [17], where just changed the dimension of the parameter ϑ.

Let us introduce the normalized likelihood ratio function

$$Z_n(u) = \frac{V(\vartheta_0 + \frac{u}{\sqrt{n}}, X^n)}{V(\vartheta_0, X^n)}, \qquad u \in U_n, \tag{1.15}$$

where the set

$$U_n = \left\{ u : \quad \vartheta = \vartheta_0 + \frac{u}{\sqrt{n}} \in \Theta \right\}.$$

Note that for the BE we have the relations (below $\vartheta_u = \vartheta_0 + \frac{u}{\sqrt{n}}$)

$$\begin{aligned}
\tilde{\vartheta}_n &= \frac{\int_{U_n} \left(\vartheta_0 + \frac{u}{\sqrt{n}}\right) p\left(\vartheta_u\right) V\left(\vartheta_0 + \frac{u}{\sqrt{n}}, X^n\right) \mathrm{d}u}{\int_{U_n} p\left(\vartheta_u\right) V\left(\vartheta_0 + \frac{u}{\sqrt{n}}, X^n\right) \mathrm{d}u} \\
&= \vartheta_0 + \frac{1}{\sqrt{n}} \frac{\int_{U_n} u\, p\left(\vartheta_u\right) \frac{V\left(\vartheta_0 + \frac{u}{\sqrt{n}}, X^n\right)}{V(\vartheta_0, X^n)} \mathrm{d}u}{\int_{U_n} p\left(\vartheta_u\right) \frac{V\left(\vartheta_0 + \frac{u}{\sqrt{n}}, X^n\right)}{V(\vartheta_0, X^n)} \mathrm{d}u} \\
&= \vartheta_0 + \frac{1}{\sqrt{n}} \frac{\int_{U_n} u\, p\left(\vartheta_u\right) Z_n(u) \, \mathrm{d}u}{\int_{U_n} p\left(\vartheta_u\right) Z_n(u) \, \mathrm{d}u}.
\end{aligned}$$

Hence

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right) = \frac{\int_{U_n} u\, p\left(\vartheta_u\right) Z_n(u) \, \mathrm{d}u}{\int_{U_n} p\left(\vartheta_u\right) Z_n(u) \, \mathrm{d}u}.$$

Let $n \to \infty$, then from continuity of the function $p\left(\vartheta_u\right)$ at the point ϑ_0 it follows that $p\left(\vartheta_u\right) \to p\left(\vartheta_0\right)$. Suppose that we already proved that the process $Z_n\left(u\right), u \in U_n$ converges in distribution to the random process

$$Z\left(u\right) = \exp\left\{ \left\langle u, \Delta\left(\vartheta_0\right)\right\rangle - \frac{1}{2} u^{\mathbb{T}} \mathbb{I}(\vartheta_0) u \right\}, \qquad u \in R^d,$$

where the vector $\Delta(\vartheta_0) \sim \mathcal{N}(0, \mathbb{I}(\vartheta_0))$ (we denote $\langle \cdot, \cdot \rangle$ the scalar product in R^d and $\|\cdot\|$ is the related norm) and that we can justify the convergence of the corresponding integrals.

Then

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right) \Longrightarrow \frac{\int_{R^d} u\, Z(u)\, \mathrm{d}u}{\int_{R^d} Z(u)\, \mathrm{d}u}.$$

Further, denote $v = \mathbb{I}(\vartheta_0)^{1/2} u$ and $\tilde{\Delta} = \mathbb{I}(\vartheta_0)^{-1/2} \Delta(\vartheta_0)$.

Using these notations we can write

$$\langle u, \Delta(\vartheta_0)\rangle - \frac{1}{2} u^{\mathbb{T}} \mathbb{I}(\vartheta_0) u = \langle v, \tilde{\Delta}\rangle - \frac{|v|^2}{2} = -\frac{1}{2}\left\|v - \tilde{\Delta}\right\|^2 + \frac{1}{2}\|\Delta\|^2$$

and

$$\begin{aligned}\int_{R^d} u\, Z(u)\, \mathrm{d}u &= \mathbb{I}(\vartheta_0)^{-1} \Delta(\vartheta_0) \int_{R^d} Z(u)\, \mathrm{d}u \\ &\quad + \mathbb{I}(\vartheta_0)^{-1/2} \int_{R^d} \left(v - \tilde{\Delta}\right) \exp\left\{-\frac{1}{2}\left\|v - \tilde{\Delta}\right\|^2\right\} \mathrm{d}v\, e^{-\frac{\|\tilde{\Delta}\|^2}{2}} \\ &= \mathbb{I}(\vartheta_0)^{-1} \Delta(\vartheta_0) \int_{R^d} Z(u)\, \mathrm{d}u.\end{aligned}$$

Hence

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right) \Longrightarrow \mathbb{I}(\vartheta_0)^{-1} \Delta(\vartheta_0) \quad \sim \quad \mathcal{N}\left(0, \mathbb{I}(\vartheta_0)^{-1}\right)$$

and we obtain the asymptotic normality of the estimator. This convergence together with the uniform integrability of M_n provides the convergence of moments.

For example, if we show that for some $p > 0$

$$\sup_{\vartheta_0 \in \Theta} \mathbf{E}_{\vartheta_0} \left\|\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right)\right\|^{p'} < C. \tag{1.16}$$

Then for any $p < p'$ we have the convergence of the moments

$$\lim_{n\to\infty} \mathbf{E}_{\vartheta_0} \left\|\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta_0\right)\right\|^{p} \longrightarrow \mathbf{E}_{\vartheta_0} \left\|\mathbb{I}(\vartheta_0)^{-1} \Delta(\vartheta_0)\right\|^{p}.$$

To verify the convergence

$$M_n = \frac{\int_{U_n} u\, p(\vartheta_u)\, Z_n(u)\, \mathrm{d}u}{\int_{U_n} p(\vartheta_u)\, Z_n(u)\, \mathrm{d}u} \Longrightarrow \frac{\int_{R^d} u\, Z(u)\, \mathrm{d}u}{\int_{R^d} Z(u)\, \mathrm{d}u}$$

we have to prove

$$\int_{U_n} u\, p\left(\vartheta_u\right) Z_n(u)\, \mathrm{d}u \Longrightarrow p\left(\vartheta_0\right) \int_{R^d} u\, Z(u)\, \mathrm{d}u,$$
$$\int_{U_n} p\left(\vartheta_u\right) Z_n(u)\, \mathrm{d}u \Longrightarrow p\left(\vartheta_0\right) \int_{R^d} Z(u)\, \mathrm{d}u$$

and for some $m \geq p' + 2$ the estimate

$$\mathbf{P}_{\vartheta_0}\left\{\|M_n\| \geq \frac{1}{N^m}\right\} \leq \frac{C}{N^m}. \tag{1.17}$$

As it was shown by Ibragimov and Hasminskii [3], the conditions $H1 - H4$ allow to justify these limits and the estimate (1.17).

The next lemma provides the verification of $H1$.

Lemma 1 *Let the conditions of regularity be fulfilled, then we have the representation*

$$Z_n\left(u\right) = \exp\left\{\left\langle u, \Delta_n\left(\vartheta_0, X^n\right)\right\rangle - \frac{1}{2} u^{\mathbb{T}} \mathbb{I}(\vartheta_0) u + r_n\right\}, \tag{1.18}$$

where $r_n \to 0$ *and*

$$\Delta_n\left(\vartheta_0, X^n\right) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\ell}\left(\vartheta_0, X_{j-1}, X_j\right) \Longrightarrow \Delta\left(\vartheta_0\right) \quad \sim \quad \mathcal{N}\left(0, \mathbb{I}(\vartheta_0)\right). \tag{1.19}$$

Proof. The representation (1.18)-(1.19) is known as *local asymptotical normality* (LAN) of the family of measures corresponding to this model of observations (see, e.g., [14], [15], [6], [17]).

Below we use the expansions $\ln\left(1 + x\right) = x - \frac{x^2}{2} + O\left(x^3\right)$ and

$$\pi(\vartheta_0 + \frac{u}{\sqrt{n}}, X_{j-1}, X_j) = \pi(\vartheta_0, X_{j-1}, X_j) + \frac{1}{\sqrt{n}} \langle u, \dot{\pi}(\tilde{\vartheta}, X_{j-1}, X_j)\rangle.$$

We have

$$\begin{aligned}
\ln Z_n(u) &= \sum_{j=1}^{n} \ln \frac{\pi(\vartheta_0 + \frac{u}{\sqrt{n}}, X_{j-1}, X_j)}{\pi(\vartheta_0, X_{j-1}, X_j)} \\
&= \sum_{j=1}^{n} \ln \left(\frac{\pi(\vartheta_0, X_{j-1}, X_j) + \frac{1}{\sqrt{n}} \langle u, \dot{\pi}(\tilde{\vartheta}, X_{j-1}, X_j) \rangle}{\pi(\vartheta_0, X_{j-1}, X_j)} \right) \\
&= \sum_{j=1}^{n} \ln \left(1 + \frac{\langle u, \dot{\pi}(\tilde{\vartheta}, X_{j-1}, X_j) \rangle}{\pi(\vartheta_0, X_{j-1}, X_j)} \frac{1}{\sqrt{n}} \right) = \langle u, \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\dot{\pi}(\tilde{\vartheta}, X_{j-1}, X_j)}{\pi(\vartheta_0, X_{j-1}, X_j)} \rangle \\
&\quad - u^{\mathbb{T}} \sum_{j=1}^{n} \frac{\dot{\pi}(\tilde{\vartheta}, X_{j-1}, X_j) \dot{\pi}(\tilde{\vartheta}, X_{j-1}, X_j)}{2n\pi(\vartheta_0, X_{j-1}, X_j)^{\mathbb{T}}} u + o(1) \\
&= \langle u, \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\ell}(\vartheta_0, X_{j-1}, X_j) \rangle \\
&\quad - u^{\mathbb{T}} \frac{1}{2n} \sum_{j=1}^{n} \dot{\ell}(\vartheta_0, X_{j-1}, X_j) \dot{\ell}(\vartheta_0, X_{j-1}, X_j)^{\mathbb{T}} u + o(1) \\
&= \langle u, \Delta_n (\vartheta_0, X^n) \rangle - u^{\mathbb{T}} \frac{1}{2n} \sum_{j=1}^{n} \dot{\ell}(\vartheta_0, X_{j-1}, X_j) \dot{\ell}(\vartheta_0, X_{j-1}, X_j)^{\mathbb{T}} u + o(1).
\end{aligned}$$

Here and in the sequence $o(1)$ means convergence to zero in probability. By the central limit theorem

$$\Delta_n (\vartheta_0, X^n) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \dot{\ell}(\vartheta_0, X_{j-1}, X_j) \Longrightarrow \Delta (\vartheta_0) \sim \mathcal{N}(0, \mathbb{I}(\vartheta_0))$$

and by the law of large numbers

$$\frac{1}{n} \sum_{j=1}^{n} \dot{\ell}(\vartheta_0, X_{j-1}, X_j) \dot{\ell}(\vartheta_0, X_{j-1}, X_j)^{\mathbb{T}} \longrightarrow \mathbb{I}(\vartheta_0).$$

Lemma 2 *Let the conditions of regularity be fulfilled. Then there exists the constant $C > 0$ such that*

$$\mathbf{E}_{\vartheta_0} |Z_n(u_2)^{1/2} - Z_n(u_1)^{1/2}|^2 \leq C |u_2 - u_1|^2.$$

Proof. Following Lemma 3.1.1 in [3] we can write (below $u_s = u_1 + s(u_2 - u_1)$) that

$$Z_n(u_2)^{1/2} - Z_n(u_1)^{1/2}$$
$$= \frac{1}{2\sqrt{n}} \int_0^1 Z_n^{1/2}(u_s) \sum_{j=1}^n \langle \dot{\ell}\left(\vartheta_0 + \frac{u_s}{\sqrt{n}}, X_{j-1}, X_j\right), (u_2 - u_1)\rangle ds.$$

Hence

$$\mathbf{E}_{\vartheta_0} |Z_n(u_2)^{1/2} - Z_n(u_1)^{1/2}|^2$$
$$\leq \frac{1}{4n} \int_0^1 \mathbf{E}_{\vartheta_0} Z_n(u_s) \left(\sum_{j=1}^n \langle \dot{\ell}\left(\vartheta_0 + \frac{u_s}{\sqrt{n}}, X_{j-1}, X_j\right), (u_2 - u_1)\rangle \right)^2 ds$$
$$\leq \frac{1}{4n} \int_0^1 \mathbf{E}_{\vartheta_{u_s}} \left(\sum_{j=1}^n \langle \dot{\ell}\left(\vartheta_0 + \frac{u_s}{\sqrt{n}}, X_{j-1}, X_j\right), (u_2 - u_1)\rangle \right)^2 ds$$
$$= \sum_{i,j=1}^n \frac{(u_2 - u_1)}{4n} \int_0^1 \mathbf{E}_{\vartheta_{u_s}} \dot{\ell}(\vartheta_{u_s}, X_{j-1}, X_j) \dot{\ell}(\vartheta_{u_s}, X_{i-1}, X_i)^{\mathbb{T}} ds (u_2 - u_1)$$
$$= \frac{1}{4} (u_2 - u_1)^{\mathbb{T}} \int_0^1 \mathbb{I}(\vartheta_{u_s}) \mathrm{d}s (u_2 - u_1) \leq C \|u_2 - u_1\|^2,$$

because the information matrix satisfies (1.4).

Lemma 3 *Let the condition of regularity be fulfilled. Then for any $q > 0$ there exist constant B_q such that*

$$\mathbf{E}_{\vartheta_0} (Z_n(u)^{1/2}) \leq \frac{B_q}{|u|^q}. \tag{1.20}$$

Proof. First we follow the proof of the Theorem 1 in [17] and write

$$\mathbf{E}_{\vartheta_0, x} Z_n(u)^{\frac{1}{2}} \leq \exp\left(-c|u|^2\right) + \frac{C_k}{(1+n)^k}.$$

We have $u = \sqrt{n}(\vartheta - \vartheta_0)$. Hence

$$|u| \leq \sqrt{n} D(\Theta), \qquad D(\Theta) = \sup_{\vartheta_1, \vartheta_2 \in \Theta} |\vartheta_1 - \vartheta_2|.$$

Further

$$n \geq \frac{|u|^2}{D(\Theta)^{2k}}, \qquad \frac{C_k}{(1+n)^k} \leq \frac{C_k D(\Theta)^2}{|u|^{2k}} = \frac{B_q}{|u|^q}.$$

with some constant $B_q > 0$.

Therefore the conditions H1-H4 of the Theorem 3.2.1 in [3] are fulfilled and according to this theorem the BE has the mentioned in Theorem 1 properties.

1.4 Method of moments

The properties of the estimators constructed by the method of moments are well-known. Nevertheless we recall here the conditions of the asymptotic normality and convergence of moments because in the forthcoming work these estimators will be used as preliminary in the construction of the asymptotically efficient estimators using the one-step and two-step MLE procedures [9].

Let $q(x)$ be a such vector-function that the function

$$m(\vartheta) = \mathbf{E}_\vartheta \; q(X) = \int g(x) \; \pi^*(\vartheta, x) \; \mathrm{d}x$$

admits a unique solution for all $\vartheta \in \Theta$ of the equation

$$m(\vartheta) = t, \qquad t \in R^d.$$

To have this property we introduce the condition of
Identifiability: *For any* $\nu > 0$ *and any* ϑ_0 *we have*

$$\kappa(\nu) = \inf_{|\vartheta - \vartheta_0| > \nu} |m(\vartheta) - m(\vartheta_0)| > 0.$$

Note that if this condition is not fulfilled then the consistent estimation of the parameter ϑ with such $q(\cdot)$ is impossible and we have to seek another function $q(\cdot)$. Indeed, if for some $\nu > 0$ we have $\kappa(\nu) = 0$ then it follows that there exists at least one ϑ_1 such that $m(\vartheta_1) = m(\vartheta_0)$ and $\vartheta_1 \neq \vartheta_0$. For example, the condition of Identifiability is fulfilled in one-dimensional case $d = 1$ if the function $m(\vartheta)$ is strictly monotone.

Moreover we suppose that this solution ϑ can be written as $\vartheta = h(t)$, where $h(\cdot) \in R^d$ is some smooth vector-function.

Recall that $\pi^*(\vartheta, x)$ is the one-dimensional invariant density and we denote its distribution function as $\Pi^*(\vartheta, x)$.

The solution $\vartheta = m^{-1}(t) = h(t)$ we write as

$$\vartheta = h\left(\int q\left(x\right) \mathrm{d}\Pi^{*}\left(\vartheta, x\right)\right).$$

Let us introduce the empirical distribution function

$$\hat{\Pi}_{n}^{*}\left(x\right) = \frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\{X_j < x\}}.$$

Now the estimator of the method of moments (EMM) $\bar{\vartheta}_n$ we obtain by substitution $\hat{\Pi}_{n}^{*}\left(\cdot\right)$ on the place of $\Pi^{*}\left(\vartheta, x\right)$. This yields

$$\bar{\vartheta}_n = h\left(\int q\left(x\right) \mathrm{d}\hat{\Pi}_{n}^{*}\left(x\right)\right) = h\left(\frac{1}{n}\sum_{j=1}^{n} q\left(X_j\right)\right).$$

The calculation of EMM can be simpler than that of the MLE or BE, but its limit covariance $\mathbb{B}\left(\vartheta\right)$ in

$$\sqrt{n}\left(\bar{\vartheta}_n - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{B}\left(\vartheta\right)\right)$$

is usually greater than that of the MLE: $\mathbb{B}\left(\vartheta\right) \geq \mathbb{I}\left(\vartheta\right)^{-1}$. This means that $\mathbb{B}\left(\vartheta\right) - \mathbb{I}\left(\vartheta\right)^{-1}$ is positive definite.

Remind that by the central limit theorem for stationary strongly mixing sequences $\xi_j = q\left(X_j\right) - \mathbf{E}_{\vartheta} q\left(X_j\right)$ we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \xi_j \Longrightarrow \mathcal{N}\left(0, \mathbb{D}\left(\vartheta\right)\right),$$

where the covariance matrix

$$\mathbb{D}\left(\vartheta\right) = \mathbf{E}_{\vartheta}\, \xi_0 \xi_0^{\mathbb{T}} + 2\sum_{j=1}^{\infty} \mathbf{E}_{\vartheta}\, \xi_0 \xi_j^{\mathbb{T}}.$$

Introduce as well the matrix

$$M\left(\vartheta\right) = \dot{m}\left(\vartheta\right) \dot{m}\left(\vartheta\right)^{\mathbb{T}}$$

and suppose that it is uniformly nondegenerate

$$\inf_{\vartheta \in \Theta} \inf_{|\lambda|=1} \lambda^{\mathbb{T}} M\left(\vartheta\right) \lambda > 0. \tag{1.21}$$

To verify the convergence of moments we need one technical lemma.

Lemma 4 *Under regularity conditions for any $L > 0$ and $m >$ there exist a constant $C > 0$ such that*

$$\mathbf{P}_{\vartheta_0}\left\{\left|\sqrt{n}\left(\bar{\vartheta}_n - \vartheta_0\right)\right| > L\right\} \leq \frac{C}{L^m}. \tag{1.22}$$

Proof. We can write

$$\begin{aligned}
&\mathbf{P}_{\vartheta_0}\left\{\left|\sqrt{n}\left(\bar{\vartheta}_n - \vartheta_0\right)\right| > L\right\} \\
&\quad = \mathbf{P}_{\vartheta_0}\left\{\inf_{|\vartheta-\vartheta_0|<Ln^{-1/2}} |\bar{m}_n - m\left(\vartheta\right)| > \inf_{|\vartheta-\vartheta_0|\geq Ln^{-1/2}} |\bar{m}_n - m\left(\vartheta\right)|\right\} \\
&\quad \leq \mathbf{P}_{\vartheta_0}\left\{\inf_{|\vartheta-\vartheta_0|<Ln^{-1/2}} \left(|\bar{m}_n - m\left(\vartheta_0\right)| + |m\left(\vartheta_0\right) - m\left(\vartheta\right)|\right)\right. \\
&\qquad\qquad \left. > \inf_{|\vartheta-\vartheta_0|\geq Ln^{-1/2}} \left(|m\left(\vartheta_0\right) - m\left(\vartheta\right)| - |\bar{m}_n - m\left(\vartheta_0\right)|\right)\right\} \\
&\quad \leq \mathbf{P}_{\vartheta_0}\left\{|\bar{m}_n - m\left(\vartheta_0\right)| > \kappa\left(Ln^{-1/2}\right)\right\}.
\end{aligned}$$

For the function $|m\left(\vartheta\right) - m\left(\vartheta_0\right)|$ we can obtain two estimates. The first local estimate is

$$\begin{aligned}
|m\left(\vartheta\right) - m\left(\vartheta_0\right)|^2 = \left|\langle \dot{m}(\tilde{\vartheta}), (\vartheta - \vartheta_0)\rangle\right|^2 &= \left(\vartheta - \vartheta_0\right)^{\mathbb{T}} \dot{m}(\tilde{\vartheta})\dot{m}(\tilde{\vartheta})^{\mathbb{T}}\left(\vartheta - \vartheta_0\right) \\
&\geq \kappa\left|\vartheta - \vartheta_0\right|^2
\end{aligned}$$

for $|\vartheta - \vartheta_0| \leq \nu$ sufficiently small ν. Remind that the matrix $M(\tilde{\vartheta}) = \dot{m}(\tilde{\vartheta})\dot{m}(\tilde{\vartheta})^{\mathbb{T}}$ is uniformly nondegenerate (1.21).

Outside of the circle $|\vartheta - \vartheta_0| \leq \nu$ we have

$$|m\left(\vartheta\right) - m\left(\vartheta_0\right)| \geq \kappa\left(\nu\right) \geq \kappa\left(\nu\right)\frac{|\vartheta - \vartheta_0|}{D\left(\vartheta\right)},$$

where we denoted $D\left(\vartheta\right)$ the diameter of the set Θ.

Therefore we have the following estimate for all $\vartheta \in \Theta$

$$|m\left(\vartheta\right) - m\left(\vartheta_0\right)| \geq \bar{\kappa}\left|\vartheta - \vartheta_0\right| \tag{1.23}$$

with some constant $\bar{\kappa} > 0$.

This estimate allows us to write

$$\begin{aligned}
\mathbf{P}_{\vartheta_0}\left\{|\bar{m}_n - m\left(\vartheta_0\right)| > \kappa\left(Ln^{-1/2}\right)\right\} &\leq \mathbf{P}_{\vartheta_0}\left\{|\bar{m}_n - m\left(\vartheta_0\right)| > \bar{\kappa}\, Ln^{-1/2}\right\} \\
&\leq \bar{\kappa}^{-m} n^{m/2} \mathbf{E}_{\vartheta_0}\left|\bar{m}_n - m\left(\vartheta_0\right)\right|^m L^{-m} \leq \frac{C}{L^m},
\end{aligned}$$

because we suppose that the moments converge (1.3).

Let us denote

$$\mathbb{B}(\vartheta) = h'(m(\vartheta))^{\mathbb{T}} \mathbb{D}(\vartheta)^{\mathbb{T}} \mathbb{D}(\vartheta) h'(m(\vartheta)).$$

Theorem 2 *Let the conditions of regularity be fulfilled, then the EMM $\bar{\vartheta}_n$ is consistent, asymptotically normal*

$$\sqrt{n}\left(\bar{\vartheta}_n - \vartheta_0\right) \Longrightarrow \mathcal{N}(0, \mathbb{B}(\vartheta_0)) \tag{1.24}$$

and we have the convergence of moments: for all $p > 0$

$$\lim_{n\to\infty} n^{\frac{p}{2}} \mathbf{E}_{\vartheta_0} \left|\bar{\vartheta}_n - \vartheta_0\right|^p = \mathbf{E}_{\vartheta_0} |\zeta(\vartheta_0)|^p, \tag{1.25}$$

where $\zeta \sim \mathcal{N}(0, \mathbb{B}(\vartheta_0))$.

Proof. The consistency follows immediately from the estimate (1.22) as follows: for any $\nu > 0$ we have

$$\mathbf{P}_{\vartheta_0}\left\{\left|\bar{\vartheta}_n - \vartheta_0\right| > \nu\right\} \leq \frac{C}{\nu^m n^{m/2}} \longrightarrow 0.$$

From the law of large numbers and continuity of the function $h(\cdot)$ we obtain the consistency of this estimator

$$\bar{m}_n = \frac{1}{n}\sum_{j=1}^{n} q(X_j) \longrightarrow \mathbf{E}_\vartheta\, q(X), \qquad \bar{\vartheta}_n \longrightarrow h(m(\vartheta)) = \vartheta.$$

To verify the asymptotic normality of this estimator first note that by the central limit theorem for Markov sequences we have

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} \left[q(X_j) - \mathbf{E}_\vartheta\, q(X)\right] \Longrightarrow \mathcal{N}(0, \mathbb{D}(\vartheta)).$$

Hence, if we suppose that the function $h(\cdot)$ is continuously differentiable at the point $m(\vartheta)$ then by Taylor's expansion we obtain

$$\begin{aligned}\bar{\vartheta}_n &= h\left(m(\vartheta) + \frac{1}{\sqrt{n}}\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[q(X_j) - \mathbf{E}_\vartheta\, q(X)\right]\right) = h\left(m(\vartheta) + \frac{\eta_n}{\sqrt{n}}\right) \\ &= h(m(\vartheta)) + h'(m(\vartheta))\frac{\eta_n}{\sqrt{n}} + o(1) = \vartheta + h'(m(\vartheta))\frac{\eta_n}{\sqrt{n}} + o(1).\end{aligned}$$

Here $h'\left(m\left(\vartheta\right)\right)$ is $d \times d$ matrix.

Therefore

$$\sqrt{n}\left(\bar{\vartheta}_n - \vartheta\right) = h'\left(m\left(\vartheta\right)\right)\eta_n + o\left(1\right)$$

and

$$\sqrt{n}\left(\bar{\vartheta}_n - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{B}\left(\vartheta\right)\right).$$

To prove the convergence of moments (1.25) it is sufficient to verify that the family of random variables is uniformly integrable: for any $p > 0$ there exists a constant $C > 0$ such that

$$n^{p/2}\mathbf{E}_{\vartheta_0}\left|\bar{\vartheta}_n - \vartheta_0\right|^p \leq C. \tag{1.26}$$

Let us denote

$$G_n\left(u\right) = \mathbf{P}_{\vartheta_0}\left\{\sqrt{n}\left|\bar{\vartheta}_n - \vartheta_0\right| < u\right\}$$

and take $m = p + 1$. We have

$$\begin{aligned}
n^{p/2}\mathbf{E}_{\vartheta_0}\left|\bar{\vartheta}_n - \vartheta_0\right|^p &= \int_0^\infty u^p \,\mathrm{d}G_n\left(u\right) \\
&= \int_0^L u^p \,\mathrm{d}G_n\left(u\right) - \int_L^\infty u^p \,\mathrm{d}\left[1 - G_n\left(u\right)\right] \\
&\leq C_1 + L^p + p\int_L^\infty u^{p-1}\left[1 - G_n\left(u\right)\right]\,\mathrm{d}u \\
&\leq C_1 + L^p + \int_L^\infty u^{p-1}\,\frac{C}{u^m}\,\mathrm{d}u \leq C_1 + L^p + \frac{C}{L} \leq C.
\end{aligned}$$

Hence the condition (1.26) is fulfilled and we have the convergence of moments.

1.5 Examples

In this section we introduce three different examples and we studied the behavior of the parameter estimators in each case. In the next chapter examples 2 and 3 will be used to illustrate the one and two-step MLE-processes.

1.5.1 Example 1

This first example was formulated in the work [17].

$$X_{n+1} = X_n - (2+\vartheta)\frac{\operatorname{sgn}(X_n)\,\ln(1+|X_n|)}{1+|X_n|} + \varepsilon_{n+1},\ \vartheta \in (-1,+1), \tag{1.27}$$

$$\varepsilon_n \sim \mathcal{N}(0,1), i.i.d.$$

This example is one for which the general theorem 2 (MLE asymptotic minimax efficiency) of [17] works.

We can not obtain analytical expression for the invariant density that is why we use the kernel type density estimation. Let us estimate the invariant density of observations of Markov sequence $X^n = (X_1, X_2, \ldots, X_n)$ with the help of kernel density estimation which is a non-parametric method.

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{j=1}^{n} K(\frac{X_j - x}{h_n}),$$

where h_n is the step and K is the symmetric kernel that satisfy the following conditions:

$$\begin{aligned} K(x) &\geq 0, \\ \int_R K(x)\,dx &= 1, \\ \int_R xK(x)\,dx &= 0, \\ \int_R x^2K(x)\,dx &< \infty. \end{aligned}$$

For the estimation of $\hat{f}_n(x)$ we need to chose the smoothing kernel to be used. This may be, for example, gaussian, rectangular, triangular, epanechnikov, etc.

The classic example of K is the gaussian kernel:

$$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\,\mathbb{1}_{[-\infty;+\infty]}.$$

We use the function density to perform kernel density estimation in R. We can show that for the large samples there is no difference of the type of smoothing

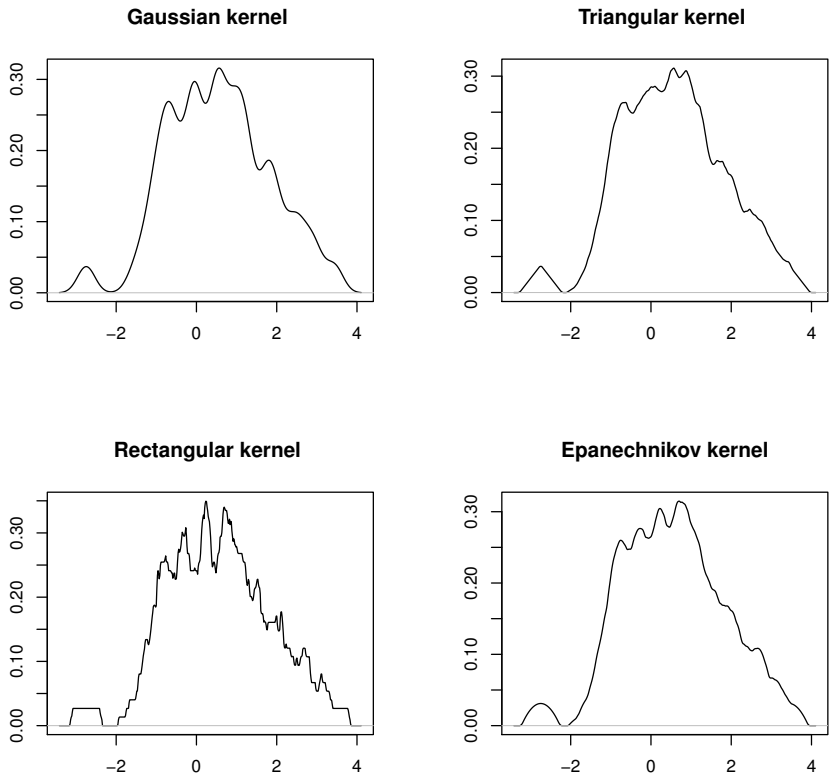

Figure 1.1: Kernel density estimation for small samples

kernel to be used. However, for the small size samples the choice of type of the kernel could be more important. On the Figure 1.1 we draw the density estimation for the sample of $n = 100$ observations.

And the Figure 1.2 represent the density estimation for the sample of $n = 10000$ observations. Here we see that all type of smoothing kernel give the same result. Therefore, below we will use the Gaussian kernel by default.

Next on the Figure 1.3 we present the kernel density estimation (with Gaussian smoothing) for different $\vartheta \in (-1, +1)$ in the case of large samples.

The next step is to calculate the maximum-likelihood estimator (MLE).

Let us denote from equation (1.27)

$$A(X_{j-1}) = \frac{\operatorname{sgn}(X_{j-1})\,\ln(1 + |X_{j-1}|)}{1 + |X_{j-1}|} \tag{1.28}$$

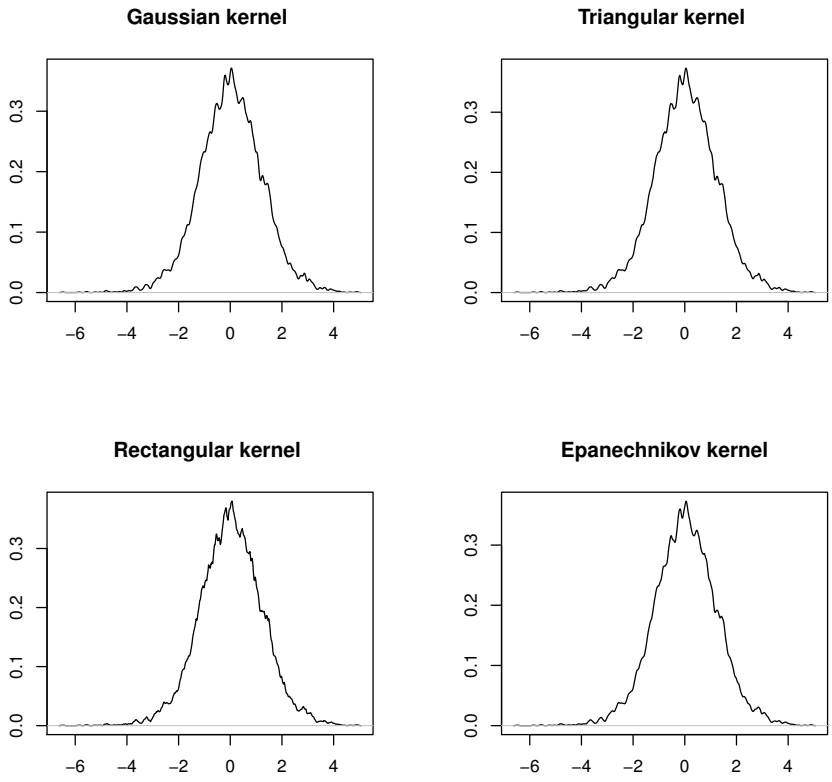

Figure 1.2: Kernel density estimation for large samples

the part of (1.27) that does not depend on parameter ϑ.

Considering (1.28) we can write the equation (1.27) in the following form

$$X_j = X_{j-1} - (2+\vartheta)A(X_{j-1}) + \varepsilon_j,\ \vartheta \in (-1,+1). \tag{1.29}$$

Let us consider the likelihood function for our example

$$V(\vartheta, X^n) = \prod_{j=1}^{n} f(\vartheta, X_{j-1}, X_j), \qquad \vartheta \in (-1,+1),$$

where

$$f(\vartheta, X_{j-1}, X_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[X_j - (X_{j-1} - (2+\vartheta)A(X_{j-1})]^2}$$

is the density function of X_n.

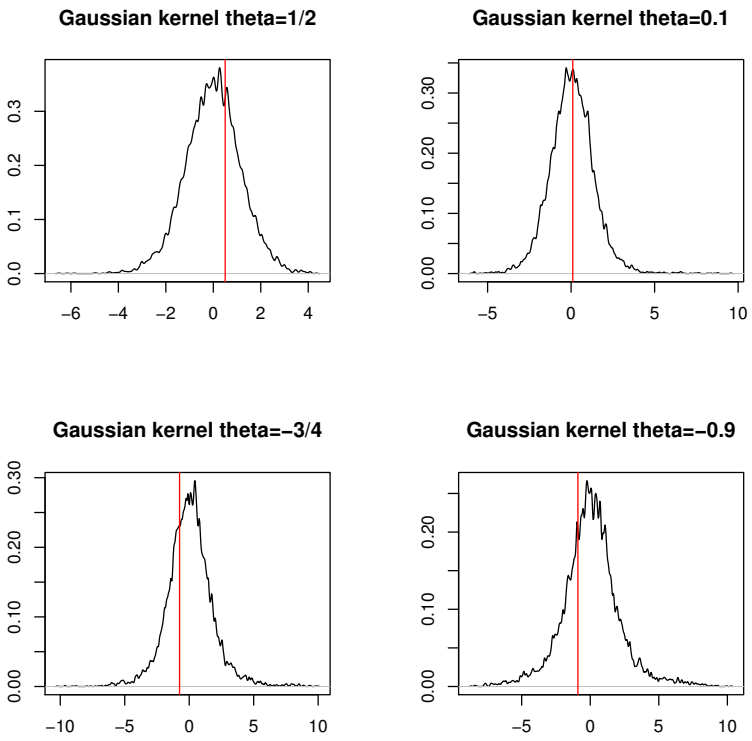

Figure 1.3: Kernel density estimation for different theta

And the log-likelihood ratio function is

$$L_n(\vartheta, X^n) = \sum_{j=1}^{n} -\frac{1}{2}[X_j - (X_{j-1} - (2+\vartheta)A(X_{j-1})]^2$$
$$= \sum_{j=1}^{n} \ell(\vartheta, X_{j-1}, X_j).$$

The partial derivative of the log-likelihood ratio function with respect to ϑ is

$$\begin{aligned}\frac{\partial L}{\partial \vartheta} &= \sum_{j=1}^{n} \dot{\ell}(\vartheta, X_{j-1}, X_j) \\ &= \frac{1}{2}\, 2 \sum_{j=1}^{n} [X_j - (X_{j-1} - (2+\vartheta)A(X_{j-1}))]A(X_{j-1}) \\ &= \sum_{j=1}^{n} [X_j - X_{j-1} - 2A(X_{j-1}) - \vartheta A(X_{j-1})]A(X_{j-1}).\end{aligned}$$

We need to find ϑ for which

$$\frac{\partial L}{\partial \vartheta} = 0.$$

$$\sum_{j=1}^{n} [X_j - X_{j-1} - 2A(X_{j-1})]A(X_{j-1}) = \vartheta \sum_{j=1}^{n} A(X_{j-1})^2.$$

So the maximum-likelihood estimator

$$\hat{\vartheta}_{MLE} = \frac{\sum_{j=1}^{n} [X_j - X_{j-1} - 2A(X_{j-1})]A(X_{j-1})}{\sum_{j=1}^{n} A(X_{j-1})^2}.$$

With the help of R we draw the distribution of log-likelihood ratio function $L_n(\vartheta, X^n)$ and we find its maximum. We compare the results for $n = 1000$ and $n = 100000$.

On the Figures 1.4 and 1.5 we present the distribution of log-likelihood ratio function $L_n(\vartheta, X^n)$ where $n = 1000$ and $n = 100000$. The red line is the true value of the parameter and the dotted line is the maximum-likelihood estimator for this parameter.

For the large samples we observe that the estimation of the parameter $\hat{\vartheta}_{MLE}$ is very close to it true value.

1.5.2 Example 2

In our first example taking from the work of Varakin A.B., Veretennikov A.Yu. (2002) [17] we have shown that $\hat{\vartheta}_{MLE}$ has an explicit expression. That's why we

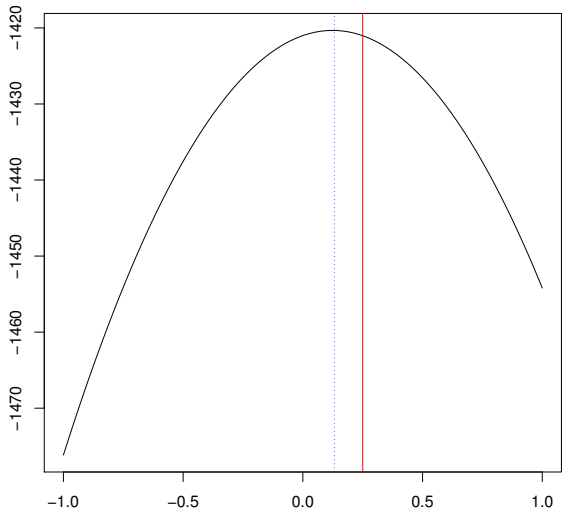

Figure 1.4: Log-likelihood function for 1 000 observations and $\vartheta = 0.25$

would not use this example to illustrate the construction of the one-step MLE-process.

Therefore let us introduce another example for which the problem of MLE calculation is evident.

We propose to study the Markov sequence $X^n = (X_1, X_2, \ldots, X_n)$ defined as

$$X_{n+1} = \frac{(X_n)^2}{1 + \vartheta \mid X_n \mid} + \varepsilon_{n+1},\ \vartheta \in (2, 5), \tag{1.30}$$

$$\varepsilon_n \sim \mathcal{N}(0, 1), i.i.d.$$

This example is one for which the general theorem 2 (MLE asymptotic minimax efficiency) of [17] works.

Let us estimate the invariant density of observations of Markov sequence $X^n = (X_1, X_2, \ldots, X_n)$ with the help of kernel density estimation we have already done in previous example.

$$\hat{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K(\frac{X_j - x}{h_n}),$$

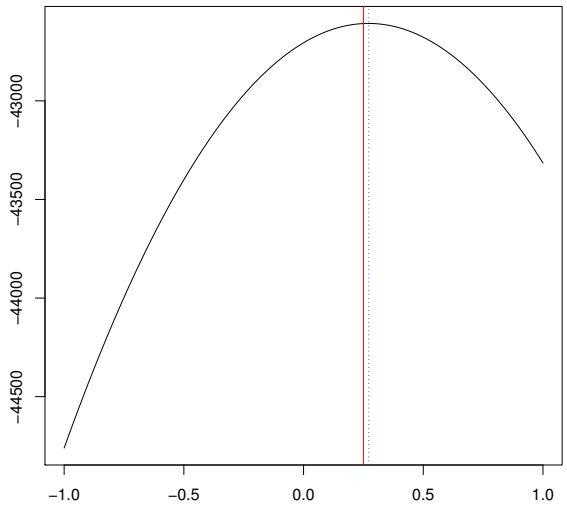

Figure 1.5: Log-likelihood function for 100 000 observations and $\vartheta = 0.25$

where h_n is the step and K is the symmetric kernel that satisfy the conditions indicated in example 1.

We use the function density to perform kernel density estimation in R. We can show that for the large samples there is no difference of the type of smoothing kernel to be used.

On the Figure 1.6 we draw the density estimation for the sample of $n = 10000$ observations.

As in the previous example, the next step is to calculate the maximum-likelihood estimator (MLE).

Let $f(\vartheta, X_{j-1}, X_j)$ be the density function of Markov sequence X^n defined in (1.30)

$$f(\vartheta, X_{j-1}, X_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[X_j - \frac{(X_{j-1})^2}{1+\vartheta|X_{j-1}|}\right]^2}. \tag{1.31}$$

Let us consider the likelihood function for our example

$$V(\vartheta, X^n) = \prod_{j=1}^{n} f(\vartheta, X_{j-1}, X_j), \qquad \vartheta \in (2, 5).$$

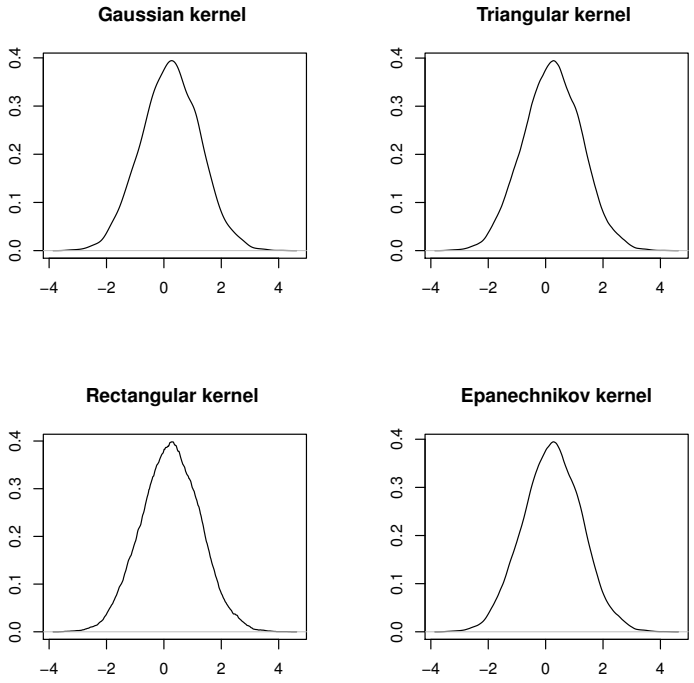

Figure 1.6: Density for Markov sequence

And the log-likelihood ratio function is

$$L_n(\vartheta, X^n) = \sum_{j=1}^{n} \left(-\frac{1}{2} \ln 2\pi - \frac{1}{2} \left[X_j - \frac{(X_{j-1})^2}{1 + \vartheta \mid X_{j-1} \mid} \right]^2 \right)$$
$$= \sum_{j=1}^{n} \ell(\vartheta, X_{j-1}, X_j).$$

The partial derivative of the log-likelihood ratio function with respect to ϑ is

$$\frac{\partial L}{\partial \vartheta} = \sum_{j=1}^{n} \dot{\ell}(\vartheta, X_{j-1}, X_j)$$
$$= \sum_{j=1}^{n} \frac{\mid X_{j-1} \mid^3}{(1 + \vartheta \mid X_{j-1} \mid)^2} \left(-X_j + \frac{(X_{j-1})^2}{1 + \vartheta \mid X_{j-1} \mid} \right).$$

We need to find ϑ for which

$$\frac{\partial L}{\partial \vartheta} = 0.$$

With the help of R we draw the distribution of log-likelihood ratio function $L_n(\vartheta, X^n)$ and we find its maximum. We compare the results for $n = 1000$ and $n = 10000$.

On the Figure 1.7 we present the distribution of log-likelihood ratio function $L_n(\vartheta, X^n)$ where $n = 1000$. The red line is the true value of the parameter and the dotted line is the maximum-likelihood estimator for this parameter.

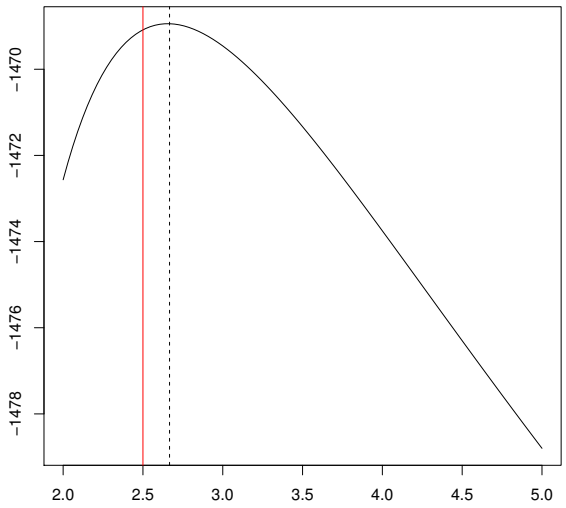

Figure 1.7: Log-likelihood function for 1 000 observations and $\vartheta = 2.5$

On the Figure 1.8 we present the distribution of log-likelihood ratio function $L_n(\vartheta, X^n)$ where $n = 10000$. The red line is the true value of the parameter and the dotted line is the maximum-likelihood estimator for this parameter.

For the large samples we observe that the estimation of the parameter $\hat{\vartheta}_{MLE}$ is very close to it true value.

In the following section we will use this example for constructing the one-step MLE-process.

Let us consider another example for which there is no explicit MLE expression.

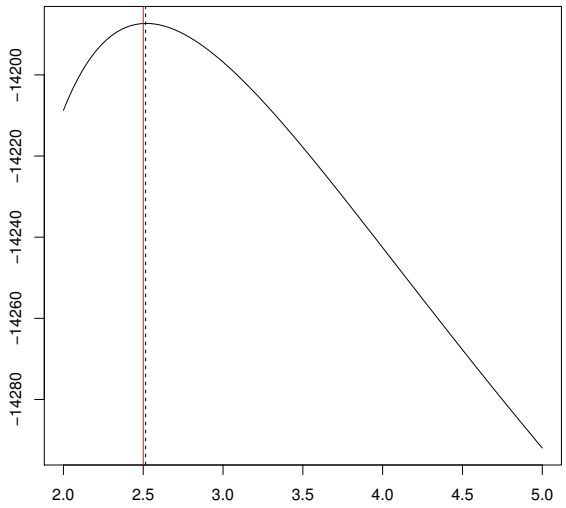

Figure 1.8: Log-likelihood function for 10 000 observations and $\vartheta = 2.5$

1.5.3 Example 3

Let us consider the following model of observations

$$X_{j+1} = X_j + 3\frac{\vartheta - X_j}{1 + (X_j - \vartheta)^2} + \varepsilon_{j+1}, \quad j = 0, 1, \ldots, n-1,$$

where $(\varepsilon_j)_{j\geq 1}$ are i.i.d. standard Gaussian random variables. The unknown parameter $\vartheta \in \Theta = (-1, 1)$. The initial value X_0 is given.

The time series $(X_j)_{j\geq 1}$ has ergodic properties with the density of invariant distribution presented on the 1.9. The vertical line represents the true value of parameter.

It is easy to see that ϑ is the shift parameter. Indeed, we have

$$X_{j+1} - \vartheta = X_j - \vartheta + 3\frac{\vartheta - X_j}{1 + (X_j - \vartheta)^2} + \varepsilon_{j+1}$$

and if we put $Y_j = X_j - \vartheta$ then

$$Y_{j+1} = Y_j - 3\frac{Y_j}{1 + Y_j^2} + \varepsilon_{j+1}.$$

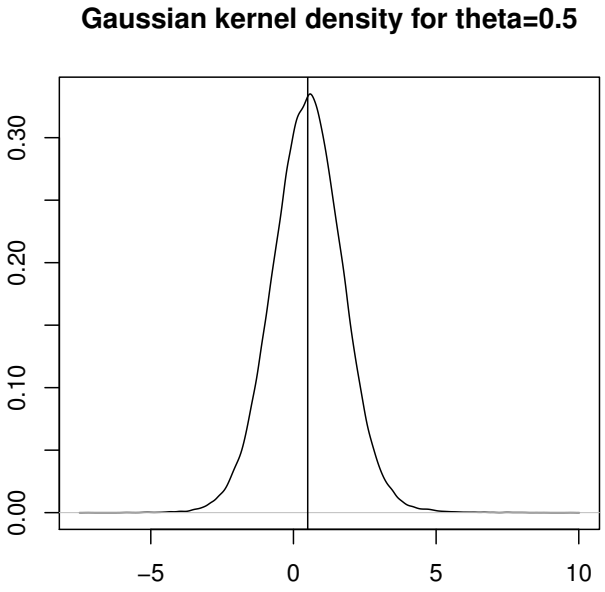

Figure 1.9: Gaussian kernel density for 100 000 observations

The invariant distribution of Y_j does not depend on ϑ.

The Fisher information does not depend on ϑ because ϑ is a shift parameter, i.e., $\mathrm{I}(\vartheta) = \mathrm{I}$.

We have

$$\ell(\vartheta, x, x') = \ln \pi(\vartheta, x, x') = -\frac{1}{2}\left(x' - x - 3\,\frac{\vartheta - x}{1 + (\vartheta - x)^2}\right)^2 - \frac{1}{2}\ln 2\pi$$

and

$$\dot{\ell}(\vartheta, x, x') = 3\left(x' - x - 3\frac{\vartheta - x}{1 + (\vartheta - x)^2}\right)\frac{1 - (\vartheta - x)^2}{(1 + (\vartheta - x)^2)^2}.$$

Therefore the Fisher information has the representation

$$\mathrm{I} = 9\,\mathbf{E}_\vartheta\left(\left(X_{j+1} - X_j - 3\frac{\vartheta - X_j}{1 + (\vartheta - X_j)^2}\right)^2 \times \left(\frac{1 - (\vartheta - X_j)^2}{(1 + (\vartheta - X_j)^2)^2}\right)^2\right)$$

$$= 9\,\mathbf{E}_\vartheta\,(\varepsilon_{j+1})^2 \times \mathbf{E}_\vartheta\left(\frac{1 - (\vartheta - X_j)^2}{(1 + (\vartheta - X_j)^2)^2}\right)^2 = 9\,\mathbf{E}_{\vartheta=0}\left(\frac{1 - \zeta_0^2}{(1 + \zeta_0^2)^2}\right)^2.$$

It can be approximated as follows

$$I_n = \frac{1}{n}\sum_{j=1}^{n}\left(\frac{1-X_j^2}{\left(1+X_j^2\right)^2}\right)^2.$$

Numerical simulations with $n = 1000$ gives us the value $I_{1000} = 2.134$ and ${\mathrm{I}_{1000}}^{-1} = 0.4686$.

Before studying the limit variances of estimator of the method of moments and Bayes estimator let us represent on the Figure 1.10 the three estimators of parameter ϑ for $n = 1000$.

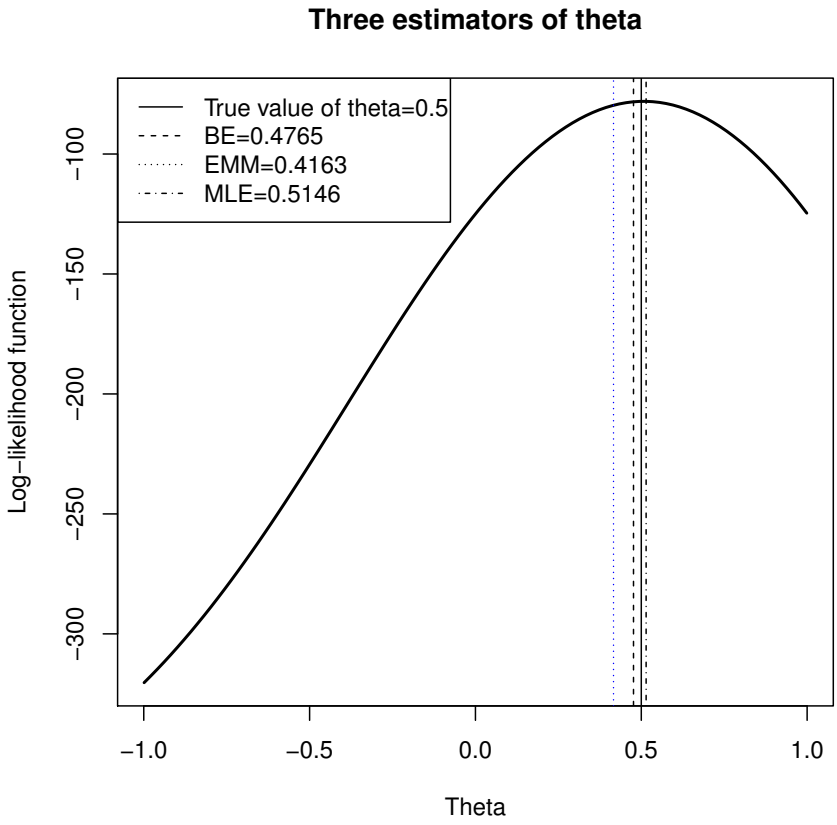

Figure 1.10: Three estimators of theta

Here we used continuous line for drawing the true value of our parameter which is $\vartheta_0 = 0.5$, dash line for Bayes estimator $\tilde{\vartheta} = 0.4765$, dotted line for estimator of the method of moments $\bar{\vartheta} = 0.4163$ and dash-dotted line for maximum-likelihood

estimator $\hat{\vartheta} = 0.5146$. Both BE and MLE are quite close to the true value ϑ_0 and the EMM gives us the worst estimation of the parameter.

Bayesian estimator.

Let us take the uniform prior distribution $p(\vartheta) = \frac{1}{2}\mathbb{1}_{\{|\vartheta|\leq 1\}}$ and calculate the Bayes estimator

$$\tilde{\vartheta}_n = \frac{\int_{-1}^{1} \vartheta\, V(\vartheta, X^n)\, \mathrm{d}\vartheta}{\int_{-1}^{1} V(\vartheta, X^n)\, \mathrm{d}\vartheta}.$$

The true value is $\vartheta = 0,5$. Denote $\tilde{u}_n = \sqrt{n}\left(\tilde{\vartheta}_n - \vartheta\right)$.

We know that

$$\sqrt{n}\left(\tilde{\vartheta}_n - \vartheta\right) \Longrightarrow \mathcal{N}\left(0, \mathrm{I}^{-1}\right).$$

Numerical simulations with $n = 1000$ repeated 1000 times give us the data $\tilde{u}_{1000,l} = \sqrt{1000}\left(\tilde{\vartheta}_{1000,l} - \vartheta\right), l = 1, \ldots, 1000$ and allow to calculate the empirical variance

$$\frac{1}{1000}\sum_{l=1}^{1000}\left(\tilde{u}_{1000,l}\right)^2 = 0.482$$

which is in accordance with the waited value $\mathrm{I}^{-1} = 0.4686$.

Estimator of the method of moments.

As ϑ is the shift parameter we can take the estimator of the method of moments as follows

$$\bar{\vartheta}_n = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

The numerical simulations of $\bar{u}_{1000,l} = \sqrt{1000}\left(\bar{\vartheta}_{1000,l} - \vartheta\right), l = 1, \ldots, 1000$ yields the value

$$B_n = \frac{1}{1000}\sum_{l=1}^{1000}\left(\bar{u}_{1000,l}\right)^2 = 1.208.$$

We see that the limit variance of the EMM is greater than that of the BE. Note that both estimators are used in the work [9] as preliminary for the construction of the asymptotically efficient estimator-processes.

Chapter 2

On Multi-step MLE-process for Markov Sequences

2.1 Introduction

This chapter is devoted to the problem of finite-dimensional parameter estimation in the case of observations of Markov sequence in the asymptotics of large samples.

Suppose we observe the process $X^n = (X_0, X_1, X_2, \ldots, X_n)$. For simplicity of exposition we take as a model of observations a nonlinear time series satisfying the relation

$$X_j = S(\vartheta, X_{j-1}) + \varepsilon_j, \quad j = 1, 2, \ldots \tag{2.1}$$

and the initial value X_0 is given too. The random variables $(\varepsilon_j)_{j\geq 1}$ are i.i.d. with some known smooth density function $g(x)$. The function $S(\vartheta, x)$ is supposed to be known and smooth with respect to ϑ.

Our goal is to construct a sequence (we say process) of estimators $\vartheta_n^\star = (\vartheta_{k,n})$, where $k = N+1, N+2, \ldots, n$ and $N \ll n$. By the first $N+1$ observations $X^N = (X_0, X_1, \ldots, X_N)$ we estimate the parameter ϑ and the obtained *preliminary* estimator $\bar{\vartheta}_N$ we use in the construction of the estimator process $\vartheta_n^\star$.

This construction is based on the modification of the well-known one-step maximum likelihood estimator (MLE) procedure introduced by Le Cam in 1956 [11]. In the proofs we follow the similar work [8] devoted to parameter estimation in the case of ergodic diffusion process.

As the initial estimator is constructed by a relatively small number of observations $N \sim n^\delta$ with $\delta < 1$ the rate of convergence of the preliminary estimator is

$\sqrt{N} \sim n^{\delta/2}$

$$\sqrt{N}\left(\bar{\vartheta}_N - \vartheta\right) \Rightarrow \mathcal{N}\left(0, \mathbb{D}\left(\vartheta\right)\right)$$

and we have *to improve* this rate up to the *optimal* $\sqrt{n}$ and *to improve* the limit variance up to the *optimal*.

Note that the idea to improve the rate of convergence of preliminary estimator using the Newton-Raphson procedure was realized by Kamatani and Uchida [5] in the different situation. They considered the problem of parameter estimation by the discrete time observations of the diffusion process in the asymptotics of the observations of *high frequency* , i.e., they supposed that the step of discretization tends to zero.

Another particularity of the presented work is the following. We propose a sequence of estimators, which can be easily calculated and the same time it has the same asymptotic properties as the asymptotically efficient MLE. This means that these estimators are asymptotically normal and that its limit variance is the inverse Fisher information matrix.

The properties of the parameter estimators for nonlinear time series and Markov sequences, of course, are well-known. Let us mention here the works by Roussas [14], Ogata and Inagaki [13], Varakin and Veretennikov [17]). More about statistical problems for time series can be found in the monographs by Veretennikov [18], Taniguchi and Kakizawa [16], Fan and Yao [4], and the references therein.

Note that we take the time series (2.1) just for simplicity of expositions. The proposed results can be generalized on the more general Markov sequences defined by their transition density if we suppose that this density satisfies to the corresponding regularity conditions.

The process $\left(X_j\right)_{j\geq 0}$ has a transition density

$$\pi\left(\vartheta, x, x'\right) = g\left(x' - S\left(\vartheta, x\right)\right).$$

It depends on the parameter θ and defines the probability of reaching the state x' after sojourning in the state x. The parameter ϑ takes its values in some open bounded set $\Theta \subset R^d$.

We suppose that the time series $\left(X_j\right)_{j\geq 1}$ is geometrically mixing, has invariant distribution with the density function $\pi^*\left(\vartheta, x\right)$ and for simplicity of exposition we put $\pi_0\left(x\right) = \pi^*\left(\vartheta, x\right)$. In this case the process $\left(X_j\right)_{j\geq 0}$ is stationary.

The construction of the one-step MLE-process in this work is done in two steps. On the first step we estimate the unknown parameter by the observations $X^N = \left(X_0, X_1, \ldots, X_N\right)$ on the *learning interval* $j \in [0, N]$. As preliminary estimator we

can take the MLE, Bayes estimator (BE), estimator of the method of moments (EMM) or any other estimator, which is consistent and asymptotically normal.

Let us recall some of them. The MLE estimator is defined as follows. Introduce the likelihood function

$$V(\vartheta, X^n) = \pi_0(X_0) \prod_{j=1}^{n} \pi(\theta, X_{j-1}, X_j), \qquad \vartheta \in \Theta, \tag{2.2}$$

where $\pi_0(x)$ is the density of the initial value X_0.

The maximum likelihood estimator we introduce as usual by the equation

$$V(\hat{\vartheta}_n, X^n) = \sup_{\vartheta \in \Theta} V(\vartheta, X^n). \tag{2.3}$$

If this equation has many solutions then we can take any of them as the MLE.

It is known that under the regularity conditions the MLE is consistent, asymptotically normal:

$$\sqrt{n}(\hat{\vartheta}_n - \vartheta) \Longrightarrow \mathcal{N}(0, \mathbb{I}(\vartheta)^{-1}), \tag{2.4}$$

where $\mathbb{I}(\vartheta)$ is the Fisher information matrix

$$\mathbb{I}(\vartheta) = \mathbf{E}^*_\vartheta \left[\dot{\ell}(\vartheta, X_0, X_1) \, \dot{\ell}(\vartheta, X_0, X_1)^{\mathbb{T}} \right],$$

where $\ell(\vartheta, x, x') = \ln \pi(\vartheta, x, x')$, the dot means the derivation w.r.t. ϑ and $\mathbb{T}$ means the transpose of a matrix. The mathematical expectation here is w.r.t. the invariant measure $\pi^*(\vartheta, \cdot)$.

As $\pi(\vartheta, x, x') = g(x' - S(\vartheta, x))$ we can write

$$\begin{aligned}
\mathbb{I}(\vartheta) &= \mathbf{E}^*_\vartheta \left[\dot{g}(X_j - S(\vartheta, X_{j-1})) \, \dot{g}(X_j - S(\vartheta, X_{j-1}))^{\mathbb{T}} \right] \\
&= \mathbf{E}^*_\vartheta \left[\frac{\dot{g}(X_j - S(\vartheta, X_{j-1})) \, \dot{g}(X_j - S(\vartheta, X_{j-1}))^{\mathbb{T}}}{g(X_j - S(\vartheta, X_{j-1}))^2} \right] \\
&= \mathbf{E} \left(\frac{g'(\varepsilon)}{g(\varepsilon)} \right)^2 \mathbf{E}^*_\vartheta \left[\dot{S}(\vartheta, \xi) \, \dot{S}(\vartheta, \xi)^{\mathbb{T}} \right] = \mathbb{I}_g \, \mathbf{E}^*_\vartheta \left[\dot{S}(\vartheta, \xi) \, \dot{S}(\vartheta, \xi)^{\mathbb{T}} \right],
\end{aligned}$$

where we used the equality $X_j - S(\vartheta, X_{j-1}) = \varepsilon_j$ and denoted

$$\mathbb{I}_g = \int \frac{g'(x)^2}{g(x)} \, \mathrm{d}x.$$

Moreover the MLE is asymptotically efficient. There are several definitions of the asymptotically efficient estimators. One of them is the following : an estimator ϑ_n^* is called *asymptotically efficient* if it satisfies the relation: for all $\vartheta_0 \in \Theta$

$$\lim_{\delta\to 0}\lim_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta}\mathbf{E}_\vartheta W\left(\sqrt{n}\left(\vartheta_n^*-\vartheta\right)\right)=\mathbf{E}W\left(\mathbb{I}(\vartheta_0)^{-1/2}\zeta\right). \tag{2.5}$$

Here $W(u), u \in R^d$ is a loss function satisfying the usual conditions. Note that it can be bounded, polynomial $W(u)=|u|^p, u \in R^d$ with $p>0$ or other (see, e.g., [3]) and ζ is a Gaussian vector $\zeta \sim \mathcal{N}(0,\mathbb{J})$, $\mathbb{J}$ is a unit $d\times d$ matrix. Remind that for all for all estimators $\bar{\vartheta}_n$ the following Hajek-Le Cam's type lower bound

$$\underline{\lim}_{\delta\to 0}\underline{\lim}_{n\to\infty}\sup_{|\vartheta-\vartheta_0|<\delta}\mathbf{E}_\vartheta W\left(\sqrt{n}\left(\bar{\vartheta}_n-\vartheta\right)\right)\geq\mathbf{E}W\left(\mathbb{I}(\vartheta_0)^{-1/2}\zeta\right) \tag{2.6}$$

holds (see, e.g. [3]). That is why (2.5) indeed defines the asymptotically efficient estimator.

Note that these properties of the MLE were established in several works. We mention here [13] and [17] (in the one-dimensional case $d=1$).

As preliminary estimator we can use as well the BE. Recall its definition and properties. Suppose that the unknown parameter $\vartheta \in \Theta$ is a random vector with the prior density $p(\vartheta), \vartheta \in \Theta$. The function $p(\cdot)$ is continuous, bounded and positive.

The Bayes estimator for the quadratic loss function has the following representation:

$$\tilde{\vartheta}_n=\frac{\int_\Theta \vartheta p(\vartheta)\,V(\vartheta,X^n)\,\mathrm{d}\vartheta}{\int_\Theta p(\vartheta)\,V(\vartheta,X^n)\,\mathrm{d}\vartheta}.$$

This estimator under regularity conditions is consistent, asymptotically normal

$$\sqrt{n}(\hat{\vartheta}_n-\vartheta_0)\Longrightarrow N(0,\mathbb{I}(\vartheta_0)^{-1}) \tag{2.7}$$

and asymptotically efficient for the polynomial loss functions. For the proof see [12].

Recall also the properties of the estimator of the method of moments. Suppose that the vector-function $q(x)\in R^d$ is such that the system of equations

$$m(\vartheta)=t,\qquad \vartheta\in\Theta,$$

where

$$m(\vartheta)=\mathbf{E}_\vartheta^* q(\xi)$$

has a unique solution $\vartheta = \vartheta(t)$.

Introduce the function $h(t)$ inverse to the function $m(\vartheta)$, i.e., $\vartheta = m^{-1}(t) = h(t)$. Then the EMM is defined as follows

$$\bar{\vartheta}_n = h\left(\frac{1}{n}\sum_{j=1}^{n} q(X_j)\right).$$

It is known that under regularity conditions this estimator is consistent, asymptotically normal

$$\sqrt{n}\left(\bar{\vartheta}_n - \vartheta\right) \Longrightarrow \mathcal{N}(0, \mathbb{C}(\vartheta)),$$

where $\mathbb{C}(\vartheta)$ is some matrix. Moreover the moments of the EMM converge too (see [12]).

In this work we use the score-function which can be calculated as follows. We introduce the log-likelihood ratio function

$$L(\vartheta, X^n) = \ln \pi^*(\vartheta, X_0) + \sum_{j=1}^{n} \ln \pi(\vartheta, X_{j-1}, X_j). \tag{2.8}$$

The normalized vector score-function

$$\begin{aligned}\Delta_n(\vartheta, X^n) &= \frac{1}{\sqrt{n}}\frac{\partial L(\vartheta, X^n)}{\partial \vartheta} \\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \frac{g'(X_j - S(\vartheta, X_{j-1}))}{g(X_j - S(\vartheta, X_{j-1}))}\dot{S}(\vartheta, X_{j-1}).\end{aligned}$$

If we denote the true value $\vartheta = \vartheta_0$, then we have

$$\Delta_n(\vartheta_0, X^n) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n} \frac{g'(\varepsilon_j)}{g(\varepsilon_j)}\ \dot{S}(\vartheta, X_{j-1}).$$

Note that $(i < j)$

$$\begin{aligned}&\mathbf{E}_\vartheta\left(\frac{g'(\varepsilon_i)}{g(\varepsilon_i)}\ \frac{g'(\varepsilon_j)}{g(\varepsilon_j)}\dot{S}(\vartheta, X_{i-1})\ \ \dot{S}(\vartheta, X_{j-1})^{\mathbb{T}}\right) \\ &\quad = \mathbf{E}_\vartheta\left(\frac{g'(\varepsilon_i)}{g(\varepsilon_i)}\ \dot{S}(\vartheta, X_{i-1})\ \mathbf{E}_\vartheta\left(\frac{g'(\varepsilon_j)}{g(\varepsilon_j)}\dot{S}(\vartheta, X_{j-1})^{\mathbb{T}}\middle|\ \mathcal{F}_{j-1}\right)\right) = 0,\end{aligned}$$

because

$$\mathbf{E}_\vartheta\left(\frac{g'(\varepsilon_j)}{g(\varepsilon_j)}\dot{S}(\vartheta, X_{j-1})^{\mathbb{T}}\middle|\ \mathcal{F}_{j-1}\right) = \mathbf{E}\left(\frac{g'(\varepsilon_j)}{g(\varepsilon_j)}\right)\mathbf{E}_\vartheta\left(\dot{S}(\vartheta, X_{j-1})^{\mathbb{T}}\middle|\ \mathcal{F}_{j-1}\right)$$

and

$$\mathbf{E}\left(\frac{g'(\varepsilon_j)}{g(\varepsilon_j)}\right) = \int_{-\infty}^{\infty} g'(x)\,\mathrm{d}x = 0.$$

Therefore by the central limit theorem

$$\Delta_n(\vartheta_0, X^n) \Longrightarrow \mathcal{N}(0, \mathbb{I}(\vartheta_0)),$$

where the Fisher information matrix

$$\mathbb{I}(\vartheta_0) = \mathbb{I}_g\,\mathbf{E}_\vartheta\left(\dot{S}(\vartheta,\xi)\;\dot{S}(\vartheta,\xi)^{\mathbb{T}}\right).$$

2.2 Main result

Suppose that we have a Markov sequence $X^n = (X_j)_{j=0,n}$ with the transition density $\pi(\cdot)$ depending on some unknown finite-dimensional parameter $\vartheta \in \Theta$. The set $\Theta \subset R^d$ is open, bounded.

Our goal is to construct on-line recurrent estimator of this parameter. Therefore we need for each j to have an estimator $\vartheta^*_{j,n}$ with *good properties*, i.e., this estimator can be easily calculated and the same time it has to be asymptotically optimal in some sense. We call such sequence of estimators $\vartheta^*_{j,n}, j = 1,\ldots,n$ *estimator-process.*

We propose a construction of such estimator in two steps. We slightly change the statement of the problem. Introduce the *learning part* $X^N = (X_0, X_1, \ldots, X_N)$ of observations $X^n = (X_0, X_1, \ldots, X_n)$, where $N = \left[n^\delta\right]$ (N is the integer part of n^δ) and the parameter $\delta < 1$ will be chosen later.

Throughout the entire paper we suppose that the following *Regularity conditions* are fulfilled.

1. *The time series* $(X_j)_{j\geq 0}$ *is geometrically mixing with the density of invariant law* $\pi^*(\vartheta, x)$ *and such that the law of large numbers*

$$\frac{1}{n}\sum_{j=1}^{n} h(X_j) \longrightarrow \mathbf{E}_\vartheta h(\xi) \equiv \int h(x)\,\pi^*(\vartheta, x)\,\mathrm{d}x$$

and the central limit theorem

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left[h\left(X_j\right)-\mathbf{E}_{\vartheta}h\left(\xi\right)\right]\Longrightarrow\mathcal{N}\left(0,\sigma\left(\vartheta\right)^2\right)$$

hold. Here we suppose that the function $h(\cdot)$ is quadratically integrable w.r.t. the invariant measure and $\sigma(\vartheta)^2<\infty$ is the corresponding limit variance.

2. *The preliminary estimator $\bar{\vartheta}_N$ is consistent and asymptotically normal*

$$\sqrt{N}\left(\bar{\vartheta}_N-\vartheta\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{B}\left(\vartheta\right)\right)$$

with some covariance matrix $\mathbb{B}(\vartheta)$.

3. *The function $S(\cdot,\cdot)\in\mathcal{C}^3_{\vartheta}$, the density $g(\cdot)>0$ and $g(\cdot)\in\mathcal{C}^3$.*

4. *The function $\ell(\vartheta,x,x')=\ln\pi(\vartheta,x,x')\in\mathcal{C}^3_{\vartheta}$ and its derivatives uniformly on ϑ are majorated by absolutely integrable functions, i.e.,*

$$\sup_{\vartheta\in\Theta}\left\|\frac{\partial^i\ell\left(\vartheta,x,x'\right)}{\partial\vartheta^i}\right\|\le R_i\left(x,x'\right),\qquad i=1,2,3,$$

where $\mathbf{E}_{\vartheta}\left|R_i\left(X_{j-1},X_j\right)\right|^2<C$.

5. *The information matrix $\mathbb{I}(\vartheta)$ is uniformly in $\vartheta\in\Theta$ non-degenerate and bounded*

$$0<\inf_{\vartheta\in\Theta}\inf_{|\lambda|=1}\lambda^{\mathbb{T}}\mathbb{I}(\vartheta)\lambda,\qquad\sup_{\vartheta\in\Theta}\sup_{|\lambda|=1}\lambda^{\mathbb{T}}\mathbb{I}(\vartheta)\lambda<\infty.\tag{2.9}$$

Here $\lambda\in R^d$.

Note that as preliminary estimator $\bar{\vartheta}_N$ we can take the MLE, the BE or the EMM. All of them have the required properties (under additional regularity conditions, which we do not mention here). The details can be found in [13], [17], [12] or any other work describing their properties. The conditions 3-4 allow us differentiate the function $\ell(\vartheta,x,x')$ with respect to ϑ and by condition 5 these derivatives have bounded moments.

We construct the one-step MLE-process $\vartheta^{\star}_{k,n},k=N+1,\ldots,n$ as follows. Introduce the variable $s\in(\tau_{\delta},1]$, where $\tau_{\delta}=n^{-1+\delta}\to0$ and put $k=[sn]$, where $[a]$ means the integer part of a. Let us write $\vartheta^{\star}_{k,n}=\vartheta^{\star}_{s,n}$ and consider the estimator-process $\vartheta^{\star}_n=\left(\vartheta^{\star}_{s,n},s\in(\tau_{\delta},1]\right)$.

Our goal is to construct an estimator process $\vartheta_n^\star$ asymptotically optimal for all $s \in (\tau_\delta, 1]$. Recall that the MLE $\hat{\vartheta}_{s,n}$ constructed by the first $k = [sn]$ observations is asymptotically efficient and for example,

$$\sqrt{sn}\left(\hat{\vartheta}_{s,n} - \theta\right) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta\right)^{-1}\right), \quad s \in [\delta, 1].$$

Note that to solve the equation

$$\sup_{\vartheta\in\Theta} V\left(\vartheta, X^{[sn]}\right) = V\left(\hat{\vartheta}_{s,n}, X^{[sn]}\right)$$

for all $s \in (\tau_\delta, 1]$ is computationally rather difficult problem, except some particular examples. Therefore it is better to seek another estimators, which have the same property to be asymptotically efficient for all $s \in (\tau_\delta, 1])$ and which can be calculated in more simple way.

We consider two different situations depending on the length of the learning interval $[0, N]$. If $N = n^\delta$ with $\frac{1}{2} < \delta < 1$ then we construct the one-step MLE-process and if we take the preliminary interval shorter, i.e., $N = n^\delta$ with $\frac{1}{4} < \delta \leq \frac{1}{2}$, then we introduce an intermediate estimator and only after that we can construct the two-step MLE-process. Therefore we consider below these two situations separately.

2.2.1 One-step maximum likelihood estimator-process

We proceed as follows. Let us fix $s \in (\tau_\delta, 1]$ and slightly modify the vector score-function

$$\Delta_k(\vartheta, X_N^k) = \frac{1}{\sqrt{k}} \sum_{j=N+1}^{k} \dot{\ell}(\vartheta, X_{j-1}, X_j),$$

where $k = [sn] \to \infty$. Introduce the one-step MLE

$$\vartheta_{s,n}^* = \bar{\vartheta}_N + \frac{1}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \Delta_k(\bar{\vartheta}_N, X_N^k).$$

Here and below for simplicity of notation this writing means that N is the integer part of n^δ.

Theorem 3 *Suppose that the conditions of regularity are fulfilled, then*

$$\sqrt{k}(\vartheta^*_{s,n} - \vartheta) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\,(\vartheta)^{-1}\right). \tag{2.10}$$

Proof. We can write

$$\begin{aligned}\sqrt{k}(\vartheta^*_{s,n} - \vartheta) &= \sqrt{k}(\overline{\vartheta}_N - \vartheta) + \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1}\Delta_k(\bar{\vartheta}_N, X_N^k)\\ &= \sqrt{k}(\overline{\vartheta}_N - \vartheta) + \mathbb{I}(\bar{\vartheta}_N)^{-1}\Delta_k(\vartheta, X_N^k)\\ &\quad + \mathbb{I}(\bar{\vartheta}_N)^{-1}\left[\Delta_k(\bar{\vartheta}, X_N^k) - \Delta_k(\vartheta, X_N^k)\right].\end{aligned}$$

We have

$$\Delta_k(\bar{\vartheta}, X_N^k) - \Delta_k(\vartheta, X_N^k) = \int_0^1 \langle(\bar{\vartheta}_N - \vartheta), \dot{\Delta}_k(\vartheta + v\left(\bar{\vartheta}_N - \vartheta\right), X_N^k)\rangle\,\mathrm{d}v.$$

Hence (below $\vartheta_v = \vartheta + v\left(\bar{\vartheta}_N - \vartheta\right)$)

$$\begin{aligned}&\sqrt{k}(\overline{\vartheta}_N - \vartheta) + \mathbb{I}(\bar{\vartheta}_N)^{-1}\left[\Delta_k(\bar{\vartheta}, X_N^k) - \Delta_k(\vartheta, X_N^k)\right]\\ &\quad = \sqrt{k}(\overline{\vartheta}_N - \vartheta)\mathbb{I}(\bar{\vartheta}_N)^{-1}\left[\mathbb{I}(\bar{\vartheta}_N) + \frac{1}{\sqrt{k}}\int_0^1 \dot{\Delta}_k(\vartheta_v, X_N^k)\,\mathrm{d}v\right].\end{aligned}$$

Further

$$\begin{aligned}&\mathbb{I}(\bar{\vartheta}_N) + \frac{1}{\sqrt{k}}\int_0^1 \dot{\Delta}_k(\vartheta_v, X_N^k)\,\mathrm{d}v = \mathbb{I}(\vartheta) + \frac{1}{\sqrt{k}}\dot{\Delta}_k(\vartheta, X_0^k) - \frac{1}{\sqrt{k}}\dot{\Delta}_k(\vartheta, X_0^{N-1})\\ &\quad + \mathbb{I}(\bar{\vartheta}_N) - \mathbb{I}(\vartheta) + \frac{1}{\sqrt{k}}\int_0^1 \left[\dot{\Delta}_k(\vartheta_v, X_N^k) - \dot{\Delta}_k(\vartheta, X_N^k)\right]\,\mathrm{d}v\\ &= \frac{1}{k}\sum_{j=1}^{k}\left[\ddot{\ell}\,(\vartheta, X_{j-1}, X_j) + \mathbb{I}(\vartheta)\right] + O\left(\frac{N}{k}\right) + O\left(n^{-\frac{\delta}{2}}\right),\end{aligned}$$

because

$$\begin{aligned}&\frac{1}{\sqrt{k}}\dot{\Delta}_k(\vartheta, X_0^{N-1}) = \frac{1}{k}\sum_{j=1}^{N-1}\ddot{\ell}\,(\vartheta, X_{j-1}, X_j) = O\left(\frac{N}{k}\right) = O\left(n^{-1+\delta}\right),\\ &\mathbb{I}(\bar{\vartheta}_N) - \mathbb{I}(\vartheta) = O\left(\bar{\vartheta}_N - \vartheta\right) = O\Big(n^{-\frac{\delta}{2}}\Big)\end{aligned}$$

and

$$\frac{1}{\sqrt{k}}\int_0^1 \left[\dot{\Delta}_k(\vartheta_v, X_N^k) - \dot{\Delta}_k(\vartheta, X_N^k)\right]\,\mathrm{d}v = O\left(\bar{\vartheta}_N - \vartheta\right) = O\Big(n^{-\frac{\delta}{2}}\Big).$$

By the central limit theorem we have

$$\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\left[\ddot{\ell}(\vartheta,X_{j-1},X_j)+\mathbb{I}(\vartheta)\right]\Longrightarrow\mathcal{N}\left(0,\mathbb{C}\left(\vartheta\right)\right)$$

with some matrix $\mathbb{C}\left(\vartheta\right)$. Remind that $\mathbf{E}_\vartheta\ddot{\ell}(\vartheta,X_{j-1},X_j)=-\mathbb{I}(\vartheta)$.

Therefore

$$\begin{aligned}\sqrt{k}(\vartheta^*_{s,n}-\vartheta)&=\mathbb{I}(\bar{\vartheta}_N)^{-1}\Delta_k(\vartheta,X_N^k)\\&\quad+n^{\frac{\delta}{2}}(\overline{\vartheta}_N-\vartheta)\left[n^{\frac{1-\delta}{2}}O\left(n^{-\frac{1}{2}}\right)+n^{\frac{1-\delta}{2}}O\left(n^{-1+\delta}\right)+n^{\frac{1-\delta}{2}}O\left(n^{-\frac{\delta}{2}}\right)\right]\\&=\mathbb{I}(\vartheta)^{-1}\Delta_k(\vartheta,X_0^k)+o\left(1\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{I}(\vartheta)^{-1}\right),\end{aligned}$$

where we used once more the central limit theorem

$$\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\dot{\ell}\left(\vartheta,X_{j-1},X_j\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{I}(\vartheta)\right).$$

Therefore the one-step MLE-process $\vartheta^\star_n=\left(\vartheta^\star_{s,n},\tau_\delta<s\leq1\right)$ for all $s\in(\tau_\delta,1]$ is asymptotically normal (2.10) with asymptotically efficient covariance matrix.

2.2.2 Two-step maximum likelihood estimator-process

The choice of the learning period of observations $N=\left[n^\delta\right]$ with $\delta\in(1/2,1)$ allows us to construct an estimator process for the values $s\in(\tau_\delta,1]$ only. It can be interesting to see if it is possible to take more short learning interval. Our goal is to show that the learning period can be $N=\left[n^\delta\right]$ with $\delta\in(1/4,1/2]$. Below we follow the construction which was already realized in [8] in the case of ergodic diffusion process.

Suppose that $N=\left[n^\delta\right]$ with $\delta\in(1/4,1/2)$. The asymptotically efficient estimator we construct in three steps. By the first N observations as before we obtain the preliminary estimator $\bar{\vartheta}_N$ which is asymptotically normal with the rate $\sqrt{N}$, i.e.,

$$n^{\frac{\delta}{2}}\left(\bar{\vartheta}_{N,1}-\vartheta\right)\Longrightarrow\mathcal{N}\left(0,\mathbb{B}\left(\vartheta\right)\right).$$

This can be the same estimator as in the preceding case. It can be, for example, the EMM, BE or MLE.

The two-step MLE-process $\vartheta_n^{\star\star} = \left(\vartheta_{k,n}^{\star\star}, k = N+1, \ldots, n\right)$ we construct as follows. Fix some $s \in (\tau_\delta, 1]$, $\tau_\delta = n^{-1+\delta}$ and introduce the second preliminary estimator-process (below $k = [sn]$)

$$\bar{\vartheta}_{k,2} = \bar{\vartheta}_{N,1} + \frac{1}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_{N,1}\right)^{-1} \Delta_k(\bar{\vartheta}_{N,1}, X^k), \tag{2.11}$$

where

$$\Delta_k(\vartheta, X^k) = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \dot{\ell}\left(\vartheta, X_{j-1}, X_j\right).$$

Then we show that the random sequence $n^{1/4+\varepsilon}\left(\bar{\vartheta}_{k,2} - \vartheta\right)$ with some $\varepsilon > 0$ is bounded in probability (tight).

Finally, using this estimator-process and the one-step procedure of the Theorem 3 we obtain asymptotically efficient estimator

$$\vartheta_k^{\star\star} = \bar{\vartheta}_{k,2} + \frac{1}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_{k,2}\right)^{-1} \Delta_k(\bar{\vartheta}_{k,2}, X^k). \tag{2.12}$$

In the next theorem we realize this program.

Theorem 4 *Suppose that the conditions of regularity are fulfilled, then the estimator $\vartheta_n^\star$ defined* (2.11) *and* (2.12) *is asymptotically normal*

$$\sqrt{k}(\vartheta_n^{\star\star} - \vartheta) \Longrightarrow \mathcal{N}\left(0, \mathbb{I}\left(\vartheta\right)^{-1}\right).$$

Proof. The only thing to proof is the tightness of the sequence of random vectors $n^{1/4+\varepsilon}\left(\bar{\vartheta}_{k,2} - \vartheta\right)$, because if it is tight, then the proof of Theorem 4 follows from the Theorem 3. Let us fix some $\varepsilon > 0$.

For the estimator-process $\bar{\vartheta}_{k,2}$ defined by (2.11) we can write

$$\begin{aligned} n^{\frac{1}{4}+\varepsilon}\left(\bar{\vartheta}_{k,2} - \vartheta\right) &= n^{\frac{1}{4}+\varepsilon}\left(\bar{\vartheta}_N - \vartheta\right) + \frac{n^{\frac{1}{4}+\varepsilon}}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \Delta_k(\bar{\vartheta}_N, X^k) \\ &= n^{\frac{1}{4}+\varepsilon}\left(\bar{\vartheta}_N - \vartheta\right) + \frac{n^{\frac{1}{4}+\varepsilon}}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \Delta_k(\vartheta, X^k) \\ &\quad + \frac{n^{\frac{1}{4}+\varepsilon}}{\sqrt{k}} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \left(\bar{\vartheta}_N - \vartheta\right) \dot{\Delta}_k(\tilde{\vartheta}_k, X^k). \end{aligned}$$

Note that $\Delta_k(\vartheta, X^k)$ is asymptotically normal and therefore

$$\frac{n^{\frac{1}{4}+\varepsilon}}{\sqrt{k}}\mathbb{I}\left(\bar{\vartheta}_N\right)^{-1}\Delta_k(\vartheta, X^k) \longrightarrow 0,$$

because $n^{\frac{1}{4}+\varepsilon}k^{-\frac{1}{2}} \to 0$. Further

$$\begin{aligned} &n^{\frac{1}{4}+\varepsilon}\left(\bar{\vartheta}_N - \vartheta\right) + \frac{n^{\frac{1}{4}+\varepsilon}}{\sqrt{k}}\mathbb{I}\left(\bar{\vartheta}_N\right)^{-1}\left(\bar{\vartheta}_N - \vartheta\right)\dot{\Delta}_k(\tilde{\vartheta}_k, X^k) \\ &\quad = n^{\frac{1}{8}+\frac{\delta}{2}}\left(\bar{\vartheta}_N - \vartheta\right)R_n, \end{aligned}$$

where

$$R_n = n^{\frac{1}{8}+\varepsilon-\frac{\delta}{2}}\left[\mathbb{J} + \mathbb{I}\left(\bar{\vartheta}_{N,1}\right)^{-1}\frac{1}{k}\sum_{j=1}^{k}\ddot{\ell}\left(\tilde{\vartheta}, X_{j-1}, X_j\right)\right].$$

We have by the law of large numbers

$$\frac{1}{k}\sum_{j=1}^{k}\ddot{\ell}\left(\vartheta, X_{j-1}, X_j\right) \longrightarrow -\mathbb{I}\left(\vartheta\right).$$

From the regularity conditions it follows that

$$\begin{aligned} &\left|\mathbb{I}\left(\bar{\vartheta}_{N,1}\right)^{-1} - \mathbb{I}\left(\vartheta\right)^{-1}\right| \leq C\left|\bar{\vartheta}_{N,1} - \vartheta\right|, \\ &\left|\dot{\Delta}_k(\tilde{\vartheta}_k, X^k) - \dot{\Delta}_k(\vartheta_k, X^k)\right| \leq C\left|\bar{\vartheta}_{N,1} - \vartheta\right|. \end{aligned}$$

Therefore we verified the tightness of the sequence $n^{\frac{1}{4}+\varepsilon}\left(\bar{\vartheta}_{k,2} - \vartheta\right)$. Now the proof of the Theorem 4 follows from the proof of the Theorem 3.

2.3 Examples

We consider below two examples. Both of them was already discussed in the previous chapter in the context of the study of the Bayesian estimators and the estimators of the method of moments. In the first example we construct the preliminary MLE and the one-step MLE-process. In the second example we construct the preliminary EMM, the second preliminary estimator-process and then the two-step MLE-process.

2.3.1 Example 1. MLE as preliminary estimator

Let us consider the problem of the construction of the one-step MLE-process in the case of observations $X^n = (X_0, X_1, \ldots, X_n)$ of the time series

$$X_j = \frac{(X_{j-1})^2}{1 + \vartheta \mid X_{j-1} \mid} + \varepsilon_j, \qquad \vartheta \in (2, 5), \tag{2.13}$$

where $(\varepsilon_j)_{j\geq 1} \sim \mathcal{N}(0, 1)$.

Note that for this model the conditions of the Theorem 2 of the work [17] can be verified. The time series has ergodic properties. The density of invariant law we estimate with the help of kernel-type estimator:

$$\hat{\pi}_n(x) = \frac{1}{nh_n} \sum_{j=1}^{n} K\left(\frac{X_j - x}{h_n}\right),$$

where the width $h_n = n^{-1/5}$ and $K(\cdot)$ is the gaussian kernel:

$$K(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}}.$$

The estimator of the invariant density in the case $n = 10^5$ and $\vartheta = 2,5.$ could be found in the previous example-section.

First we define the MLE constructed on the learning sequence $X^N = (X_0, X_1, \ldots, X_N)$. For the conditional density function $\pi(\vartheta, X_{j-1}, X_j)$ of the Markov sequence (2.13), we have the representation

$$\pi(\vartheta, X_{j-1}, X_j) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left[X_j - \frac{(X_{j-1})^2}{1+\vartheta|X_{j-1}|}\right]^2}. \tag{2.14}$$

Hence the likelihood function is

$$V(\vartheta, X^N) = \pi_0(X_0) \prod_{j=1}^{N} \pi(\vartheta, X_{j-1}, X_j), \qquad \vartheta \in (2, 5).$$

And the log-likelihood ratio function is

$$\begin{aligned} L_N(\vartheta, X^N) &= \ln \pi_0(X_0) + \sum_{j=1}^{N} \left(-\frac{1}{2} \ln 2\pi - \frac{1}{2} \left[X_j - \frac{(X_{j-1})^2}{1 + \vartheta \mid X_{j-1} \mid} \right]^2 \right) \\ &= \ln \pi_0(X_0) + \sum_{j=1}^{N} \ell(\vartheta, X_{j-1}, X_j). \end{aligned}$$

On the Figure 2.1 we present the result of simulations of this log-likelihood ratio function $L_N(\vartheta, X^n)$, where $N = n^{3/4} \sim 5623$. The continuous vertical line corresponds to the true value $\vartheta_0 = 2.5$ of the parameter and the vertical dotted line corresponds to the maximum-likelihood estimator $\hat{\vartheta}_N$.

To find the MLE we have to solve the maximum likelihood equation

$$\frac{\partial L}{\partial \vartheta} = \sum_{j=1}^{N} \dot{\ell}\left(\vartheta, X_{j-1}, X_j\right) = 0, \qquad \vartheta \in (2,5)\,,$$

which has the following form

$$\sum_{j=1}^{N} \frac{\mid X_{j-1} \mid^3}{(1+\vartheta \mid X_{j-1} \mid)^2} \left(-X_j + \frac{(X_{j-1})^2}{1+\vartheta \mid X_{j-1} \mid}\right) = 0, \qquad \vartheta \in (2,5)\,.$$

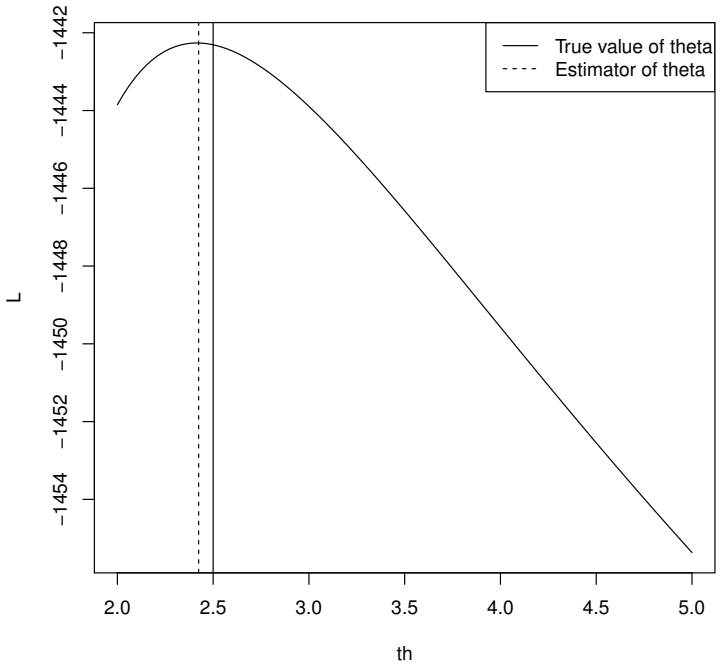

Figure 2.1: Log-likelihood function for 1 000 observations

Further we construct the one-step MLE-process $\vartheta_n^\star = \left(\vartheta_{k,n}^\star, N+1 \le k \le n\right)$ based on this preliminary estimator $\hat{\vartheta}_N$ as follows.

The normalized score-function is

$$\Delta_k(\vartheta, X^k) = \frac{1}{\sqrt{k}} \sum_{j=1}^{k} \frac{\mid X_{j-1} \mid^3}{(1 + \vartheta \mid X_{j-1} \mid)^2} \left(-X_j + \frac{(X_{j-1})^2}{1 + \vartheta \mid X_{j-1} \mid} \right),$$

where $N + 1 \leq k \leq n$. Finally the one-step MLE-process that has the following representation

$$\vartheta^{\star}_{k,n} = \hat{\vartheta}_N + \frac{1}{\mathbb{I}_k(\hat{\vartheta}_N) k} \sum_{j=1}^{k} \frac{\mid X_{j-1} \mid^3}{(1 + \hat{\vartheta}_N \mid X_{j-1} \mid)^2} \left(-X_j + \frac{(X_{j-1})^2}{1 + \hat{\vartheta}_N \mid X_{j-1} \mid} \right),$$

where $N + 1 \leq k \leq n$ and $\mathbb{I}_k(\hat{\vartheta}_N) = 0.001$ is the Fisher information calculated as follows

$$\mathbb{I}_k(\hat{\vartheta}_N) = -\frac{1}{k} \sum_{j=1}^{k} \ddot{\ell} \left(\hat{\vartheta}_N, X_{j-1}, X_j \right).$$

More detailed analysis shows that with such definition of the empirical Fisher information the main result of this work Theorem 2 is valid. Therefore the estimator-process $\vartheta^{\star}_n$ is asymptotically normal and asymptotically efficient.

The realization of the simulated one-step MLE-process for $n = 10^5$ is shown on the Figure 2.2. We can see that the initial estimator $\hat{\vartheta}_N$ is far from the true value and that the trajectory of one-step MLE-process approaches to the true value.

2.3.2 Example 2. EMM as preliminary estimator

Let us consider another example, where it will be much more easy to take the EMM as preliminary one. Our goal is to illustrate the convergence of the one-step MLE-process when the initial estimator is not asymptotically efficient. It be can, for example, the EMM which has"bad" rate and "bad" limit variance.

Introduce the time series

$$X_j = X_{j-1} + 3 \frac{\vartheta - X_{j-1}}{1 + (X_{j-1} - \vartheta)^2} + \varepsilon_j, \quad j = 1, \ldots, n, \tag{2.15}$$

where $(\varepsilon_j)_{j \geq 1}$ are i.i.d. standard Gaussian random variables. The unknown parameter $\vartheta \in \Theta = (-1, 1)$. The initial value X_0 is supposed to be given too.

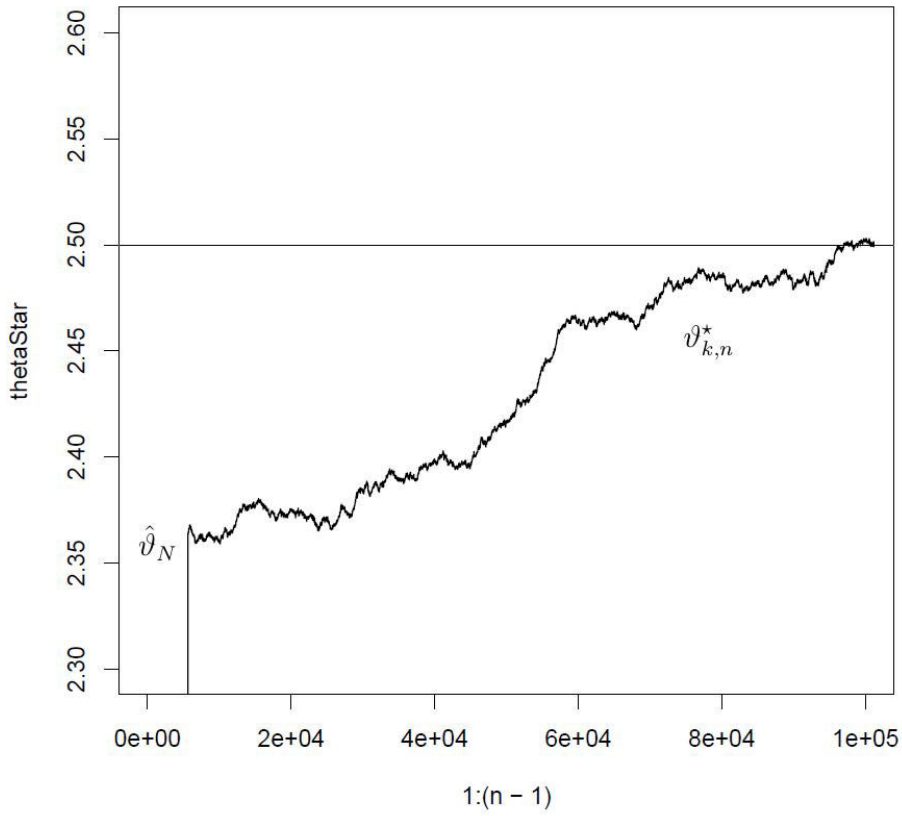

Figure 2.2: One-step MLE-process for 100 000 observations

Note that this example was already used in the work [12] and in the previous chapter to illustrate the properties of the BE and EMM.

This process has ergodic properties and its invariant density can be estimated as in the Example 1 with the help of the kernel-type estimator. The result of such estimation can be found in [12] and in the previous example-section.

In this example our goal is to two estimator-processes: one-step and two-step. Our goal is to construct the estimator-processes $\vartheta^{\star}_n$ and $\vartheta^{\star\star}_n$, which are asymptotically equivalent to the MLE and therefore are asymptotically efficient. The same time their calculation is much more simple than that of the MLE.

We start with the one-step MLE-process. As described before we construct this estimator in two steps. First we need to calculate a consistent preliminary estimator $\bar{\vartheta}_N$ by the initial observations $X_1, \ldots, X_N$, where $N = n^{\delta}$ with $\delta \in (\frac{1}{2}, 1)$. As preliminary we can take the MLE, BE or EMM. Note that the unknown parameter for this model of observations is the shift parameter and that the invariant density

function is symmetric with respect to ϑ. Hence we can take the EMM

$$\bar{\vartheta}_N = \frac{1}{N}\sum_{j=1}^{N} X_j \longrightarrow \vartheta, \qquad N = n^{3/4}.$$

Of course, the limit variance of the EMM $\bar{\vartheta}_N$ is greater than that of the BE, but this estimator is much more easier to calculate.

The score-function process is

$$\Delta_k(\vartheta, X^k) = \frac{1}{\sqrt{k}}\sum_{j=1}^{k} \dot{\ell}\left(\vartheta, X_{j-1}, X_j\right), \qquad N+1 \le k \le n,$$

where

$$\begin{aligned}\dot{\ell}\left(\vartheta, x, x'\right) = &-\frac{1}{2}\,2\left(x' - x - 3\frac{\vartheta - x}{1+(\vartheta-x)^2}\right)\times \\ &3\,\frac{1+(\vartheta-x)^2-(\vartheta-x)\,2\,(\vartheta-x)}{\left(1+(\vartheta-x)^2\right)^2} = \\ &3\left(x' - x - 3\frac{\vartheta - x}{1+(\vartheta-x)^2}\right)\,\frac{1-(\vartheta-x)^2}{\left(1+(\vartheta-x)^2\right)^2}.\end{aligned}$$

Therefore we can calculate the one-step MLE-process as follows

$$\begin{aligned}\vartheta^{\star}_{k,n} &= \bar{\vartheta}_N + \frac{1}{\mathbb{I}_k\sqrt{k}}\Delta_k(\bar{\vartheta}_N, X^k) \\ &= \bar{\vartheta}_N + \frac{3}{\mathbb{I}_k k}\sum_{j=1}^{k}\left(X_j - X_{j-1} - 3\frac{\vartheta - X_{j-1}}{1+(\vartheta-X_{j-1})^2}\right)\,\frac{1-(\vartheta-X_{j-1})^2}{\left(1+(\vartheta-X_{j-1})^2\right)^2}.\end{aligned}$$

Here $N+1 \le k \le n$ $\mathbb{I}_k$ is the empirical Fisher information. Its calculation in this example can be found in [12]. Note that $\mathbb{I}\,(\vartheta) = \mathbb{I}$ as usual with the shift parameter.

Remind that by the Theorem 2 this estimator asymptotically normal .

The simulated one-step MLE-processes are shown on the Figure 2.3 and 2.4 for $n = 1000$ and $n = 10000$.

On the Figure 2.3 the preliminary EMM $\bar{\vartheta}_N = 0.45$ that is quite close to the true value of parameter $\vartheta = 0.5$. We obtain this estimator based on the learning interval of $N = 178$ observations. And we can observe the estimator-process $\vartheta^{\star}_n = \left(\vartheta^{\star}_{k,n}, k = N+1; \ldots, n\right)$ that tends to the true value.

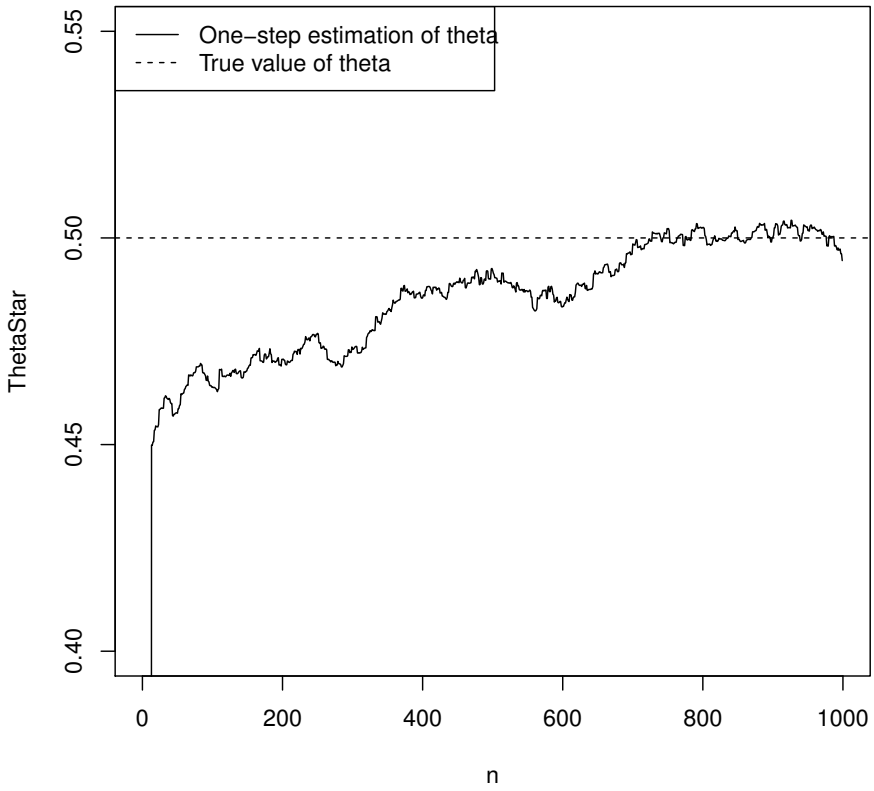

Figure 2.3: One-step MLE-process for n=1 000

On the Figure 2.4 the preliminary EMM $\bar{\vartheta}_N = 0.56$ that is quite close to the true value $\vartheta = 0.5$. We obtain this estimator based on the learning interval of $N = 1000$ observations. And we can observe the sequence of estimators $\vartheta_n^\star = \left(\vartheta_{k,n}^\star, k = N+1; \ldots, n\right)$ that tends to the true value.

Let us illustrate the two-step MLE-process. Now we take $N = n^{3/8}$.

We consider two cases: one with $n = 1000$ observations and the second with $n = 10000$ observations.

On the Figure 2.5 the preliminary EMM $\bar{\vartheta}_N = 0.4$ that is quite far from the true value $\vartheta = 0.5$. We obtain this estimator based on the learning interval of $N = 1000^{3/8} \approx 13$ observations. Then we obtain the second preliminary estimator-process $\vartheta_n^\star = \left(\vartheta_{k,n}^\star, k = N+1; \ldots, n\right)$ (continuous line) and see that it tends to the true value. The two-step MLE-process $\vartheta_n^{\star\star}$ (dashed line) is closer to the true

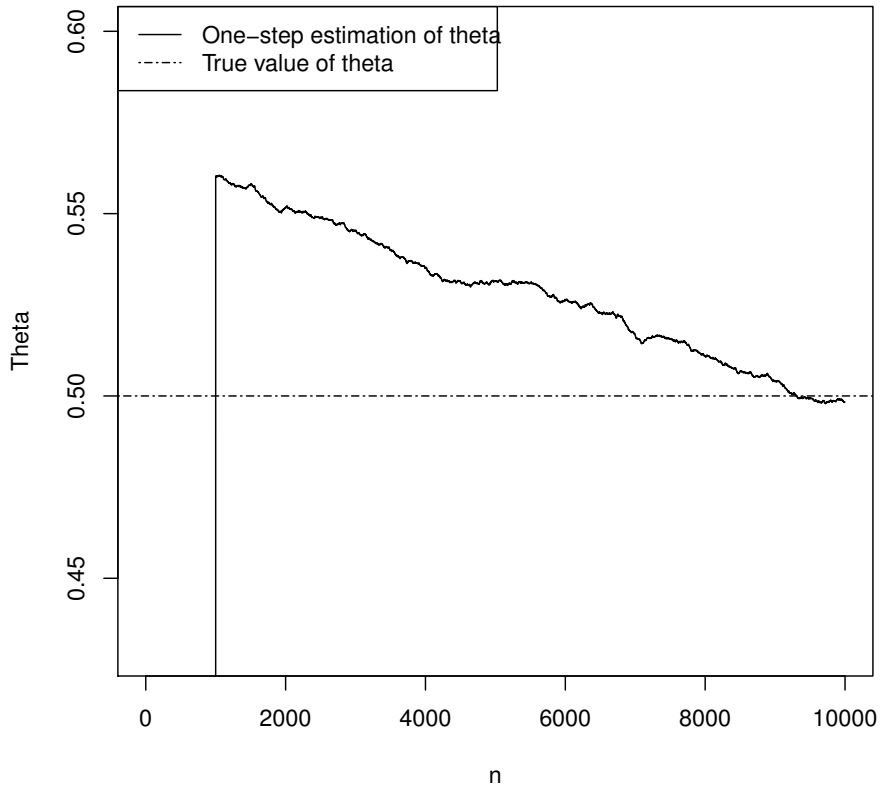

Figure 2.4: One-step MLE-process for n=10 000

value and as well tends to the true value.

On the Figure 2.6 the preliminary EMM $\bar{\vartheta}_N = 0.54$ that is quite close to the true value $\vartheta = 0.5$. We obtain this estimator based on the learning interval of $N = 10000^{3/8} \approx 32$ observations. Then we obtain the second preliminary estimator-process $\vartheta_n^\star = \left(\vartheta_{k,n}^\star, k = N+1; \ldots, n\right)$ (continuous line) and see that it tends to the true value. The two-step MLE-process $\vartheta_n^{\star\star}$ (dashed line) is closer to the true value and as well tends to the true value.

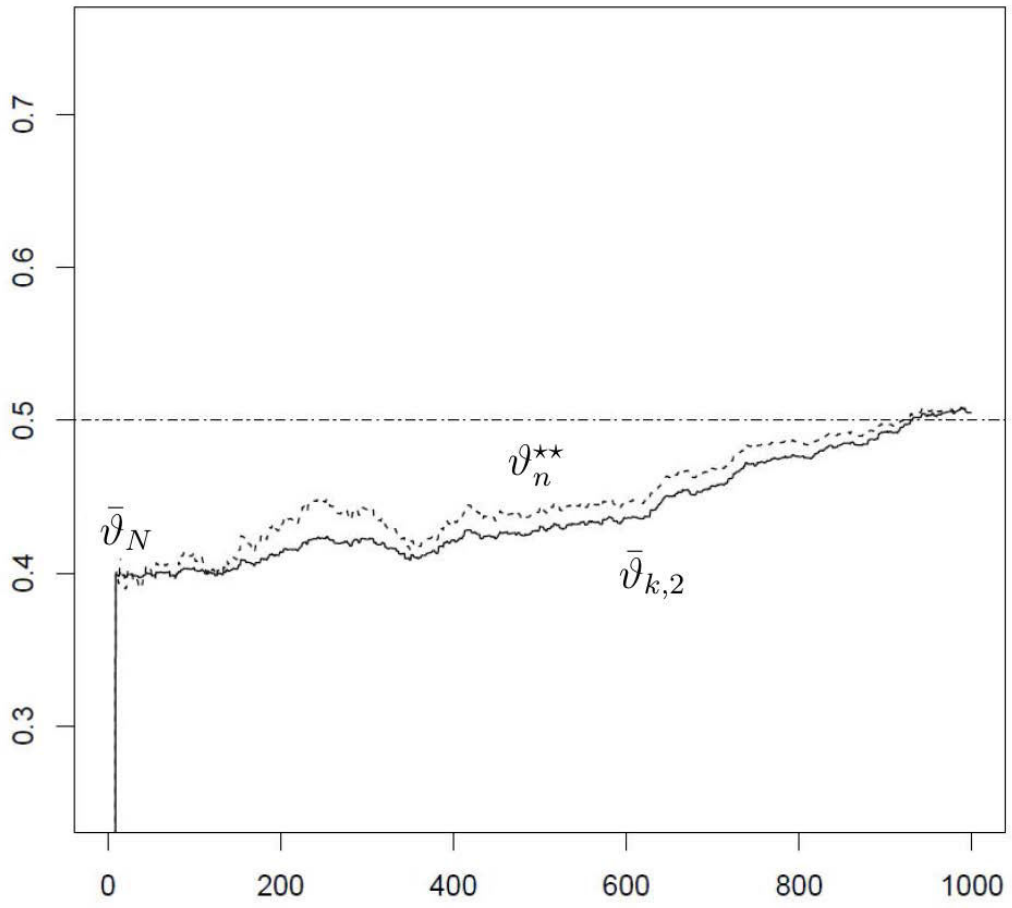

Figure 2.5: Second preliminary and two-step MLE-processes for n=1 000 and $\vartheta = 0.5$

2.4 Discussion

Two-step MLE-process allows us to estimate the parameter θ for the values k satisfying the condition $n^{1/4} < k \leq n]$. If we need a shorter learning interval, say, $\left[1, n^{\delta}\right]$ with $\delta \in (\frac{1}{8}, \frac{1}{4}]$, then we have to study the three-step MLE-process, i.e., we use a preliminary estimator $\bar{\vartheta}_N$ and two estimator-processes like (2.11).

Note that the proposed one-step MLE-process can be written in the recurrent form.

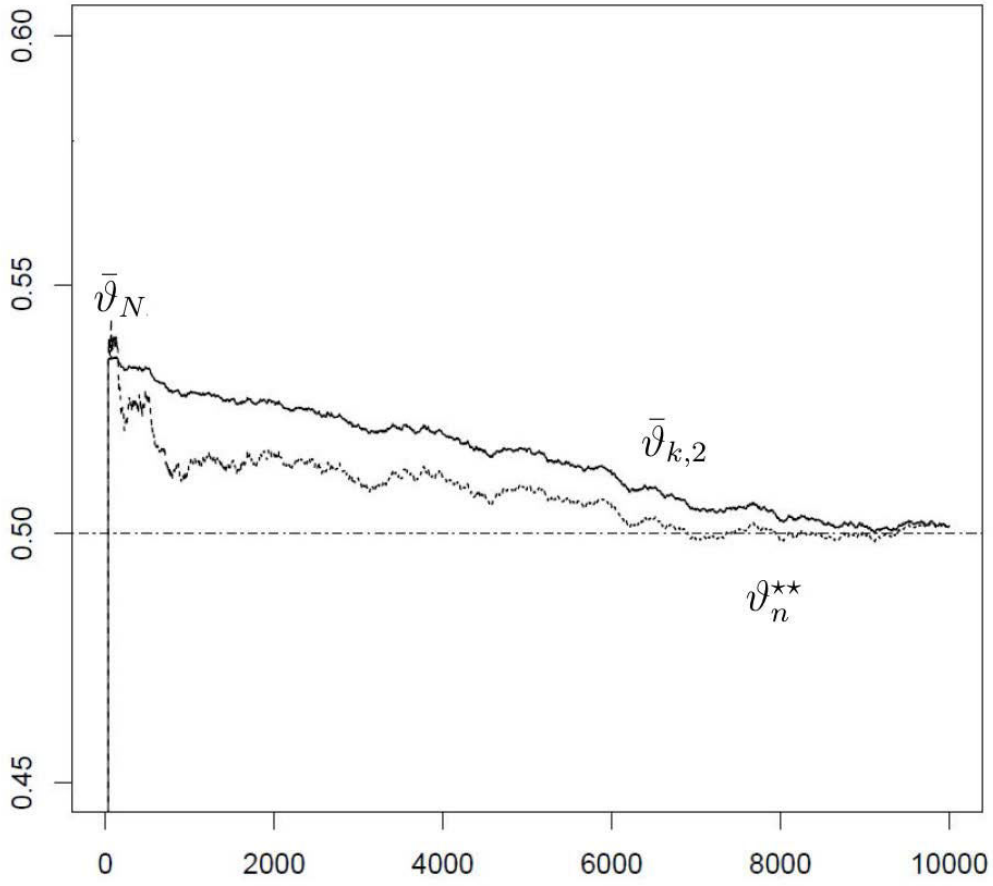

Figure 2.6: Second preliminary and two-step MLE-processes for n=10 000 and $\vartheta = 0.5$

Indeed, the estimator $\vartheta^{\star}_{k,n}$ we can represent in the following way

$$\begin{aligned}
\vartheta^{\star}_{k+1,n} &= \bar{\vartheta}_N + \frac{1}{\sqrt{k+1}} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \Delta_{k+1}\left(\bar{\vartheta}_N, X^{k+1}\right) \\
&= \bar{\vartheta}_N + \frac{1}{k+1} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \left[\sum_{j=1}^{k} \dot{\ell}\left(\bar{\vartheta}_N, X_{j-1}, X_j\right) + \dot{\ell}\left(\bar{\vartheta}_N, X_k, X_{k+1}\right)\right] \\
&= \frac{k}{k+1}\left[\bar{\vartheta}_N + \frac{1}{k} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \sum_{j=1}^{k} \dot{\ell}\left(\bar{\vartheta}_N, X_{j-1}, X_j\right)\right] + \frac{1}{k+1}\bar{\vartheta}_N \\
&\quad + \frac{1}{k+1} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \dot{\ell}\left(\bar{\vartheta}_N, X_k, X_{k+1}\right) \\
&= \frac{k}{k+1}\vartheta^{\star}_{k,n} + \frac{1}{k+1}\bar{\vartheta}_N + \frac{1}{k+1} \mathbb{I}\left(\bar{\vartheta}_N\right)^{-1} \dot{\ell}\left(\bar{\vartheta}_N, X_k, X_{k+1}\right).
\end{aligned}$$

The obtained presentation

$$\vartheta^{\star}_{k+1,n} = \frac{k}{k+1}\vartheta^{\star}_{k,n} + \frac{1}{k+1}\bar{\vartheta}_N + \frac{1}{k+1}\mathbb{I}\left(\bar{\vartheta}_N\right)^{-1}\dot{\ell}\left(\bar{\vartheta}_N, X_k, X_{k+1}\right)$$

allows us to calculate $\vartheta^{\star}_{k+1,n}$ using the values $\bar{\vartheta}_N, \vartheta^{\star}_{k,n}$ and observations X_k, X_{k+1} only.

The similar structure can be obtained for the two-step MLE-process too. Note that this is not a particular case of the well-known algorithms of stochastic approximation (see, for example the works of Duflo *Random Iterative Models*).

Chapter 3

On applications of Markov chains in health economics

3.1 Introduction

The management of critically ill patients is strongly dependent of intravascular catheters, most of them being central venous catheters (CVC) and arterial catheters. However, intravascular catheters can lead to serious infectious complications including Catheter-Related Bloodstream infections (CRBSIs) [19], [20] as they are a relevant entry door allowing microorganisms into the bloodstream [21]. CRBSI is still a frequent (1-5 episodes/1000 catheter-days) and life-threatening complication observed in critically ill patients (ICU) [22], [23], [24], [25], [26] even though in the last years different strategies have been proven to efficiently reduce its risk.

The patient's own skin flora is very likely the most important source of catheter colonization and infection for central venous catheters in place for 10 days or less, and responsible for 60% of CRBSIs [27], [28]. The skin flora microorganisms more often causing CRBSI, *Staphylococcus aureus* and *S. epidermidis*) [29], [30], [31], will regrow after skin antisepsis [32] and colonize the outer surface of the catheters [33]. CR-BSIs frequently trigger sepsis which causes a great deal of morbidity and deaths, and increases health care costs [58], [44].

A large proportion of the CR-BSIs are preventable through careful control of the factors responsible for colonization of intravascular catheters by microorganisms [34], [35]. Several interventions based on better education, training and staffing [36] or on the implementation on evidenced-based bundles of care [37], [38] were proven to be able to reducing the CRBSI rates. Also, the use of CHG antiseptic solutions for

prepping the skin before the central line insertion and bathing or cleansing the patients during their ICU stay have a positive impact in preventing CRBSIs [39], [40], [41]. Implementation of more sophisticated technologies, as catheters impregnated with antiseptics and antibiotics, also contribute to further reduce the risk of CRBSIs [42].

Antimicrobial catheter dressings are one of the available medical technologies designed to prevent skin flora re-growth and, as a consequence, to reduce the incidence of CR-BSIs. The use of a CHG-containing antimicrobial sponge at the catheter insertion site significantly reduced the CRBSI rate in Intensive Care Unit (ICU) population in France [43], even when the baseline was already low (below 2 episodes/1000 catheter-days). More recently, the clinical efficacy of a novel transparent CHG-containing catheter dressing, combining antimicrobial activity and transparency for easy visual observation of the insertion site, has been also evaluated by the same team, in a French multicenter randomized controlled trial [44]. In this study, the impact of the antimicrobial transparent dressing on reducing the CRBSI rate was also highly statistically significant.

Subsequently, two multi-state models were constructed. A homogeneous (H-MCMC) and non-homogeneous Markov Model with Markov Chain Monte Carlo simulation (NH-MCMC) models have been developed [50], based on individual patient data collected during the French RCT [77]. The methods and results of these two approaches are discussed and compared.

The aim of these works, focusing on the ICU perspective, was to evaluate the cost-effectiveness of routine use of the CHG-containing dressings in critically ill patients. The both models attempted to simulate the various observable health trajectories of ICU patients regarding the risk of acquiring CRBSIs and to evaluate all the uncertainty around the estimations of the RCT. These new models are profoundly different from all previous economical evaluations of antimicrobial dressings for intravascular access related to the prevention of CRBSI [45], [46], [47]. Those evaluations used decision-tree models representing the therapeutic choices (antimicrobial vs. non-antimicrobial dressings) and the incidence of catheter-related infections as clinical outcome.

The results of this chapter have been the subject of several publications and oral presentations. The principal article named “Cost-effectiveness analysis of a transparent antimicrobial dressing for managing central venous and arterial catheters

in intensive care units" [48] was submitted and accepted for the publication in PLOS ONE journal of science and medicine (the latest impact factor: 3.534).

Besides, several posters and presentations were closely linked with this topic. The posters were presented during the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) annuals conferences (2012-2014) and the corresponding abstracts were published in Value in Health:

1. Non-homogeneous cost-effectiveness modeling of a new CHG-dressing for preventing catheter-related bloodstream infections for patients in intensive care units [49].
2. Cost-Effectiveness of the TLC-NOSF Dressing in Venous Leg Ulcers [50].
3. Modeling cost-effectiveness of antimicrobial dressings for preventing catheter-related bloodstream infection: homogeneous vs. non-homogeneous Markov approaches [51].
4. Cost-effectiveness analysis of an antimicrobial transparent dressing for protecting central vascular accesses in critically ill patients versus standard transparent dressings in France: A comparison of two modeling approaches: Decision-Tree versus Non-Homogeneous Markov Model (NHMM) [52].

3.2 Healthcare decision-making

3.2.1 Medical interest

Catheter-related bloodstream infections (CRBSIs) are associated with attributable mortality rates of up to 11.5% and additional length of stay in the intensive care unit (ICU) of up to 12 days [53], [54]. The universally accepted method for minimizing CRBSIs is a bundle of care combining maximal sterile barrier precautions for insertion, an appropriate antiseptic solution for skin antisepsis and line access, preferential subclavian catheterization, and immediate removal of unnecessary catheters [55], [56].

Combining this catheter-care bundle with continuous quality improvement programs can decrease the CRBSI rate below 2 per 1,000 CVC-days [57], [58]. In Europe, the incidence of CRBSIs ranges from 1 to 3.1 per 1,000 patient-days [59] and according to the French surveillance network, less than one CRBSI occurred per 1,000 CVC-days in 2010 [60]. However, rates below 2 per 1,000 CVC-days are difficult to achieve in all ICUs [61], [62] and in the long term [63].

Most organisms responsible for short-term CRBSIs originate from the insertion site [64]. It was demonstrated previously that the risk of developing CRBSIs can be dramatically reduced (60% decrease) by the systematic use of a new antimicrobial transparent dressing [44] containing a Chlorhexidine Gluconate (CHG) gel even though bundles of care are appropriately followed and CRBSI level is lower than 1.5 per 1,000 catheter-days in the control group.

The purpose of this work is to evaluate the advantages of the routine use of the new CHG dressing to secure central lines of patients in ICU from a medico-economic viewpoint compared to non-antimicrobial transparent dressings, in settings where bundles of care practices are appropriately followed and where incidence of infection is already low (1.5 per 1,000 catheter-days). Both medical and economic criteria are embedded into a decision-analytic model to support the choice of the best dressing strategy from an ICU perspective.

3.2.2 Health economic context

The main question of the health economics we can formulate as follows:

- What is the additional cost required to obtain a supplementary benefits of a therapeutic intervention or treatment?
- How much should society spend to increase the life expectancy for one year?

These questions are crucial because they emphasize the fact that resources are limited. In health system it is necessary to consider the both medical and economic variables in terms of efficacy and in terms of expected cost.

The health economic evaluation provides explicit information on the costs and consequences of different interventions (health products or technologies, therapeutic sequences, screening programs, etc.) to improve decision-making and to promote efficient resource use. It is intended to facilitate the public decisions about resource allocation. The countries with a National Health Service or National health insurance generally leave the political authorities to decide on new drugs, new therapies and medical devices to be covered by the plan. It is clear that the innovative product with the proven therapeutic benefits, often associated with the reducing the risk of premature death to a certain population at risk, induces the extra cost (additional expenses) compared to existing products.

3.2.3 Cost-effectiveness analysis

The cost-effectiveness analysis (CEA) is a method of medico-economic evaluation to assess the costs and medical benefits for the various medical technologies, therapeutic sequences or concurrent clinical strategies. Generally, CEA is used to compare a therapeutic innovation with the most widely used current strategy.

The general rule in solving a problem of medical and economic decision is simple. The decision should be made taking into account two factors: in terms of efficiency and in terms of cost. Medical and budgetary approaches are opposed to each other naturally. Physicians think in terms of the expected medical service or in terms of efficiency. Managers try to minimize the expected expenditure.

The efficiency criterion is a way to bring together these two terms. In CEA the goal is

to calculate the incremental cost-effectiveness ratio (ICER). The ICER means that the additional cost generated by the new product is rewarded with medical effect.

$$ICER = \frac{additional\ costs}{additional\ benefits} = \frac{costs\ of\ innovation - cost\ of\ reference}{effect\ of\ innovation - effect\ of\ reference}.$$

The substitution or replacement of a strategy by another leads to the difference in cost and the difference in efficiency. By calculating ICERs, we can classify the different strategies with respect to another based on the efficiency criterion. Obviously, the best strategy is the one that is the least expensive and that is most effective. On the other side, the strategy is dominated by another if it is more expensive and less effective or equally effective but more expensive.

It should not be considered in any way that the medical and economic assessment is limited to simple calculation of the ICER. The real challenge for the evaluation is to incorporate into the model the proposed uncertainty concerning this calculation.

And finally, if the cost-effectiveness of a strategy is proven, the positive or negative consequences of the decision can be assessed on the budget of a payer structure. To analyze these budgetary consequences, a budget impact analysis (BIA) should be proposed by the health economists. By cons, there is no sense in BIA if the differential cost-effectiveness ratio is not favourable to the innovative strategy.

The objective of this medico-economic study is to demonstrate the advantages for patient of the routine use of CHG dressing for central lines in intensive care unit (ICU) patients, compared to non-antimicrobial dressings, related to the prevention of catheter-related bloodstream infections. The demonstration will take in consideration both medical and economic criteria and will be founded on an analytic decision model (multi-state homogeneous and non-homogeneous Markov models). The model outcomes will support the choice of the best dressing strategy based on a retro-prospective cost-effectiveness analysis (the transition probabilities are estimated from the real-life individual database).

3.3 Presentation of database

3.3.1 Data Collection

The main data source was the database assembling all patient data collected during the RCT [44]. This multicentre randomized-controlled study compared the impact of the antimicrobial CHG dressing (referred in the current dissertation as CHG dressings) and of non-antimicrobial transparent dressings (referred as non-CHG dressings) on the rate of catheter related infections.

The main objective of the RCT transposed in this cost-effectiveness analysis was to determine if the use of the new transparent CHG dressing decreased CRBSI rates. The RCT was not blinded to the investigators or ICU staff due to the obvious visual differences between the dressings, but was blinded to the microbiologists processing the skin and catheter cultures and to the committee adjudicating on the CRBSI cases. The two groups receiving different types of non-antimicrobial transparent dressings in the RCT were pooled together as "non-antimicrobial transparent dressings" for the purpose of the modeling presented in this work.

3.3.2 Study Population

The multicentre RCT [44] enrolled adult patients (>18 years) admitted to 12 French ICUs in seven universities and four general hospitals, from 31 May 2010 to 29 July 2011, and expected to require intravascular catheterization for 48 hours. Patients with known allergies to chlorhexidine or transparent dressings were excluded. Of 2,054 screened patients with at least one catheter, 1,898 could be enrolled in the study and 1,879 were assessable for the intention-to-treat analysis, for a total of 4,163 catheters and 34,339 catheter-days. Patients and catheters characteristics are reported in the next sections.

Due to a short time horizon (30 days), the patient characteristics such as age and the proportion of males/females were not incorporated in the model. However, the analysis of adjustment on covariates between the subgroups was conducted in order to ensure the comparability between two strategies (see section 3.3.4 and 3.3.5).

3.3.3 Study Catheters

In the RCT, all central venous catheters inserted at subclavian, jugular and femoral veins, as well as arterial catheters inserted at radial and femoral arteries for a given patient, were managed according to the randomized dressing assignment. Pulmonary arterial, hemodialysis, and peripherally-inserted venous catheters and catheters inserted before ICU admission were excluded from the study. All study centers followed French recommendations for catheter insertion and care, which are similar to Center for Disease Control (CDC) recommendations [65].

3.3.4 Additional ICU Length of Stay (LOS) due to CR-BSI

The analyses presented in this work were conducted on the "Global" population, comprising patients who, during their ICU stay remained alive, died or discharged from the ICU. The "global" patient is the main statistical unit of the study (see section Results of cost-effectiveness analysis).

In order to assess the impact of CR-BSI on extending ICU LOS, a subgroup analysis was performed within the "Global" population, comparing patients having developed a CR-BSI during the ICU stay with those not having developed a CR-BSI. The comparison was made through independent non-homogeneous MCMC simulations for each dressing strategy. These NH-MCMC simulations were based on observed patient data, collected during the Dressing 2 clinical study. In this study, all patients were randomly assigned to one of the two dressing strategies, what allow us to assume comparability of the groups.

Discussing further on the comparability of these two subgroups, we can present two additional adjustments.

First, a "natural" adjustment, which is linked to the main statistical unit of our modeling which is the "global" patient was considered. It means that the probability of developing a CR-BSI in the "Global population", that corresponds to the clinical trial population, follows the same plausible statistical distribution law (due to the randomization).

The statistical analysis for all confounding covariates, such as age, sex, severity (SOFA score), duration of catheterization, number of dressing change per day, shows the comparability between these subgroups (see section 3.3.5).

A further adjustment from NH-MCMC simulation was performed, considering covariates which could impact mainly the values linked to the cost-effectiveness results. The rate of catheter change and the number of additional ICU days due to CRBSI in each dressing strategy were taken into account.

3.3.5 Adjustments on covariates between the subgroups

A statistical analysis for all confounding covariates, such as age, sex, Sequential Organ Failure Assessment severity score (SOFA, a score predicting ICU mortality based on lab results and clinical data [66]), duration of catheterization, number of dressing change per day, was performed in order to demonstrate the comparability between the subgroups (see Table 1). Four subgroups of patients (CHG/CRBSI, CHG/No-CRBSI, Non-CHG/CRBSI, Non-CHG/No-CRBSI) were compared with these covariates. Mann-Whitney tests between subgroups were performed.

Table 1. Comparability of subgroups on covariates

Dressing group		CHG * Mean (std)	Non-CHG ** Mean (std)	Comparison p-value ŧ
SOFA score (severity)				
	CRBSI	7.89 (4.08)	10.29 (3.39)	0.1459
	No CRBSI	8.17 (3.76)	8.17 (3.83)	0.8737
Age (years)				
	CRBSI	58.78 (13.73)	62.57 (19.08)	0.5262
	No CRBSI	61.97 (15.71)	62.17 (16.42)	0.6043
Number of males				
	CRBSI	5 (55.56%)	12 (57.14%)	1.0000
	No CRBSI	630 (68.11%)	603 (65.97%)	0.3460
Catheterization time (days)				
	CRBSI	39.67 (22.58)	28.43 (31.56)	0.0984
	No CRBSI	11.01 (11.52)	10.92 (11.01)	0.9934
Number of dressings per day				
	CRBSI	0.59 (0.29)	0.73 (0.37)	0.2675
	No CRBSI	0.67 (0.52)	0.65 (0.58)	0.2653

* CHG group frequencies: 9 patients with CRBSI, 925 patients without CRBSI
** Non-CHG group frequencies: 21 patients with CRBSI, 914 patients without CRBSI
ł The results (p value) of Mann-Whitney tests on these covariates between subgroups show no statistically significant difference if p>0.05 (at a 0.05 level)
CHG: chlorhexidine gluconate; SOFA: Sequential Organ Failure Assessment; CRBSI: catheter-related bloodstream infection

The results of Mann-Whitney tests on these covariates between subgroups (CRBSI/No CRBSI) show no statistically significant difference at the 0.05 level.

3.3.6 Main Assumptions

1. In the cases where the "Discharge" state was reported, and a CRBSI was observed for this patient up to two days after the event, the infection was considered to occur the day of discharge from the ICU.
2. The transitional probability from the health state "Contact Dermatitis" to “Dressing Gauze and Tape” state was considered the same for both groups. By entering to the “Dressing Gauze and Tape” the patient followed probabilities of transition corresponding to the non-CHG dressings arm.
3. The cost of CRBSI is independent from the outcome (survival or death or discharge).
4. Catheter colonization with or without CRBSIs was considered as having negligible diagnosis costs and was excluded from the model for not being considered as a “health state” *per se*.
5. The costs related to replacement of a catheter suspected to be colonized (and causing CRBSI) were comprised in one of the health states including the need for a new central line. The cost per ICU day was considered as identical for each dressing group. The cost of a gauze and tape dressing is identical in both groups.
6. Health states including CRBSIs were assumed to last a single day because it was not technically possible to identify the termination of a CRBSI in the patient database. However, the costs of treating the complete episode, as well as the total costs associated with the extra length of stay due to the CRBSI were accounted on the day when the CRBSI was diagnosed.
7. With the current knowledge of publicly available data sources, there is no direct CR-BSI related risk of dying.

3.4 Medico-economic evaluation using Markov models

The use of Markov model is specially required in following cases:

- clinical trials have insufficient periods of follow-up to assess the impact of therapy in the long-term. The Markov model extrapolates from trial results in terms of transitions and rates for all comparators and estimates the cost-effectiveness of new interventions over a life-time horizon;
- to determine the influence of uncertainty surrounding input parameters. It can be achieved using several univariate and multivariate analyses (as deterministic and probabilistic sensitivity analysis where the values for parameter estimates vary within the uncertainty distributions that best reflect the nature of each specific parameter);
- when needed an accurate representation of the evaluated clinical structure by modeling repetitive events and time dependence of probabilities.

3.4.1 Markov homogeneous model

Markov models consider the patients in a discrete state of health, and the events represent the transition from one state to another. Such type of model permits a more accurate representation of the evaluated clinical structure by modeling repetitive events and time dependence of probabilities (see: time-inhomogeneous Markov Chains).

In this section we will consider the discrete-time Markov chain $X(t) = (X_0, X_1, \ldots X_n)$, where $X(t)$ is the health state in time t. In the following epidemiological applications the state space is discrete. The Markov property states that the conditional probability distribution for the system at the next step depends only upon the present state, not on the sequence of events that preceded it (it is so called memoryless property).

A random process X possesses the Markov property, and is called a Markov Chain, if

$$\mathbf{P}(X_{n+1} = j \mid X_0 = i_0,\ X_1 = i_1, \ldots,\ X_n = i_n) = \mathbf{P}(X_{n+1} = j \mid X_n = i_n)$$

depends only on i_n and j, not on any past values.

Time-homogeneous Markov chains are processes where

$$\mathbf{P}(X_{n+1} = j \mid X_n = i_n) = \mathbf{P}(X_n = j \mid X_{n-1} = i_n)$$

for all n. It implies that the probability of the transition is independent of n or the probability of the transition is the same after each step.

Markov models have limitations that must be overcome as models become more sophisticated, especially when dealing with time-dependent probabilities of transitions and different states of disease.

3.4.1.1 Study design

The adopted modeling approach complies with the guidelines of French National Authority for Health (Haute Autorité de Santé - HAS) [67]. The 30-day ICU-time homogeneous Markov model [68], [69] structure was based on observed data of a multicentre RCT [44], conducted by the Grenoble University Hospital - CHU Grenoble. The model has been programmed using Visual Basic Application with the Excel® 2007 software. This model consists of six health states described in Table 2.

Table 2. Health states for H-MCMC model defined from a multicentre randomized controlled trial

Health States	Definition
1. No CRBSI / No new CT needed	Insertion of a first catheter, no diagnosed CRBSI and no contact dermatitis
2. CRBSI / No new CT needed	CRBSI diagnosed without neither contact dermatitis nor the need for inserting a new catheter
3. Contact dermatitis	No diagnosed CRBSI, and no need for new catheter inserted but occurrence of contact dermatitis
4. Dressing Gauze and Tape	Change to an alternative dressing strategy (gauze and tape) due to contact dermatitis
5. Discharge	Patient leaves the ICU alive
6. Death	Patient dies during the ICU stay

CRBSI: Catheter-related Bloodstream Infections; CT Catheter: Central venous or radial / femoral arterial

The statistical unit of the study is the ICU patient within a time horizon of 30 days (discharged alive from the ICU, alive but still at the ICU, or deceased during the ICU stay). Patient data from the multicentre RCT [44], comparing the CHG dressing to

non-antimicrobial transparent dressings, were translated into a patient transition matrix among the different possible health states, for both the antimicrobial and non-antimicrobial dressings groups. This transitional matrix was used to perform homogeneous Markov-Chain Monte Carlo (H-MCMC) simulations [70] representing the observed daily evolution of patients in ICU. 1,000 Monte Carlo simulations of 1,000 patients were used for probabilistic sensitivity analysis and 95% confidence intervals (CI) calculations.

3.4.1.2 Model structure

Markov models consider patients in a discrete state of health [71], [72], [73], [74], and events representing the transition from one health state to another. This type of modeling can take into account iterative occurrences for each Markov state. The Markov property refers to the fact that the conditional probability distribution of future states of the health depends only upon the present state, not on the sequence of events that preceded it.

The hypothesis of this homogeneous model is that the transition probabilities do not change with time spending in ICU. This assumption is very strong and therefore the non-homogeneous model was developed (section 3.4.2).

Building up the transition matrix

The overall percentages of patient in each health state per comparator as reported in the RCT database [44] were transformed to 1-day (cycle length) transition probabilities. Each patient will be in one of the six health states described in Table 2 at each day in ICU.

3.4.1.3 Healthcare resource use and costs

Costs used within the model reflect the ICU perspective in France and consist of following components:

- drug acquisition costs,
- cost of treating the adverse events ,
- direct costs of treating the CR-BSI,
- costs due to ICU stay.

Base case input parameters considered in the cost analysis

The base case analysis is the most representative case of the real life, considering French ICU settings, and depending on expert opinions, literature, and RCTs.
The main input parameters considered in the cost analysis are the following:

- Dressing costs per day: CHG dressing which is 20 times more expensive than non-antimicrobial transparent film and 60 times costly than gauze and tape.
- Cost of treating contact dermatitis (mean/episode): catheter removal, 23.62 € [75]; four gauze and tape dressings, 0.24 €; catheter insertion, 94.87 €.
- Direct cost of treating CRBSI (mean/episode) [75]: 580.26 €.
- Cost per ICU [76]: 1,265.93 €/day.
- Additional ICU Length of stay (LOS) due to CRBSI: 9.33 days (NH-MCMC calculation).
- Cost of added ICU LOS due to CRBSI: 11,811.13 € (NH-MCMC calculation).
- Overall cost of one CRBSI (direct cost of treating one CRBSI plus cost of additional ICU LOS due to CRBSI): 12,391.40 € (calculation).

Direct costs for the treatment of CRBSIs were obtained from a micro-costing study[75]. ICU costs were based on an observational (real life) study [76] that assessed all resources consumed during a patient day in the ICU. This twenty-four hours multicentre prospective medico-economic study provides a complete overview and estimation of the actual average cost for medical and surgical ICUs in different hospital types in France: Hospitals (CH), University Hospitals (CHU) and Regional Hospitals (CHR). Twenty-two ICUs were selected randomly and all costs for 109 patients were estimated. For patients with CRBSI, an additional cost [77] due to an extra ICU length of stay (LOS) was calculated (see next section).

Main Assumptions Used for the Cost Analysis

- The cost of CR-BSI is independent from the outcome (survival or death or discharge). For the analysis the main statistical unit is the “global” (survival or death or discharge) patient;
- Catheter colonization with or without CR-BSI had no costs (after the diagnosis) or adverse outcomes (colonization has been initially excluded from the model because it’s not a “health-state”. The costs for diagnosis are negligible compared to costs related to additional LOS. The costs related to replace a

catheter suspected to be colonized (and causing CR-BSI) will be absorbed in the health-states CT new (the dressing arm with NO CHG);

- The estimated cost per ICU day at the Grenoble University Hospital is identical in each Dressing Group;
- The G+T (gauze and tape) cost for CHG-group is identical to No-CHG group.

Costs items for each Markov state

Table 3 reportes the costs included in the analysis for each Markov state.

Table 3. Costs items for each Markov state

Main costs	Detailed costs	no CR-BSI/ no CT new	CR-BSI/ no CT new	Contact dermatitis	Dressing G&T	Death	Discharge
Cost of dressings		X	X	X	X		
Cost of treating contact dermatitis				X			
	Four standard dressings			X			
	Removal of the catheter			X			
	Insertion of a new catheter			X			
Cost of treatment of CR-BSI			X				
Additional ICU -LOS			X				

Costs per Markov state per patient

The calculation of the cost for each Markov state per patient was done as follows (using the base case input parameters listed above):

- Dressing costs (including time needed per dressing, number of nurses involved, and materials used [75]) and cost per ICU day [76] were taken into account for health states 1-6;
- Cost of treating contact dermatitis [75] – (including catheter removal, four alternative dressings, and insertion of a new catheter) was taken into account only for health state 5;
- Cost of treatment of CRBSI [75] and additional ICU-LOS due to CRBSI [44], [75] were taken into account for health states 2.

The costs per patient for each health state were calculated in both CHG and No-CHG dressing and presented in Table 4.

Table 4. Costs per Markov state per patient from the base case scenario

Markov State	***Costs for 1 patient CHG, Euro 2013***	***Costs for 1 patient No-CHG, Euro 2013***
No CRBSI/noCTnew	1,268	1,266
CRBSI/noCTnew	13,659	13,657
Contact dermatitis	1,387	1,385
Dressing G+T	1,266	1,266
Discharge	0	0
Death	0	0

3.4.1.4 Results of cost-effectiveness analysis

The results presented below are from the base case scenario of the cost-effectiveness modeling. The structure follows the next principal sections:

- Base case scenario results for 1,000 H-MCMC of 1,000 patients CHG group
- Base case scenario results for 1,000 H-MCMC of 1,000 patients No-CHG group

The results showed in Tables 5 and 6 refer to the base case scenario of the cost-effectiveness modeling. The main difference between CHG group and Non-CHG

group are based on % of states 3 and 4, e.g. number of CRBSIs. A ratio of 1 to 4.67 is observed for the average number of CRBSIs between dressing groups. The number of ICU-days, the number of days before discharging and the number of days before dying are comparable in the two groups.

Table 5. Mean number of CRBSI for 1,000 patients in each dressing group – Horizon: 30-days ICU

Stat.	CHG group (1)	No-CHG group (2)	Diff. (1-2)
Number of CRBSI	1.8	8.42	-6.62

CRBSI occurred for almost 2 and 9 patients in each CHG and non-CHG groups respectively (considering 1,000 patients in each group).

The table 6 shows the cost result for the average patient in each dressing group.

Table 6. Mean Cost for 1 patient in each dressing group – Horizon: 30-days ICU

Stat.	CHG group (1)	No-CHG group (2)	Diff. Cost (1-2)
Mean cost	€ 21,748	€ 21,803	€ - 55

The statistical significance of this result will be discussed in Sensitivity analyses section.

3.4.1.5 Sensitivity analyses

Sensitivity analyses are performed to vary each parameter of the model in order to determine what levels will result in a change of preference for the therapeutic strategy. This is a way to test the boundaries of the model and identify the main parameters driving cost differences.

One-way Sensitivity Analysis

For the one-way sensitivity analysis/tornado diagram, we varied the parameters under the base-case assumptions. The resulting tornado diagram is shown in Figure 1-3. Results are most sensitive to the additional ICU LOS due to CRBSI, CHG dressing cost, dressing change schedule, CRBSI, death and discharge rate changes.

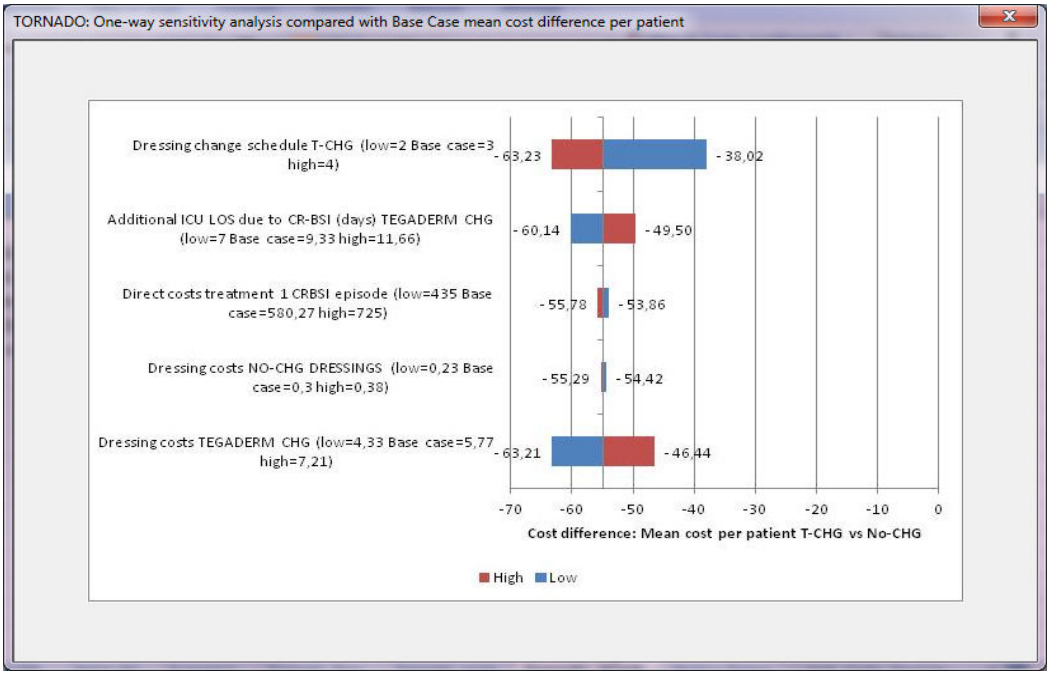

Figure 1. One-way Sensitivity Analysis: Mean cost difference per global patient between CHG and No-CHG strategies, in euros; part 1.

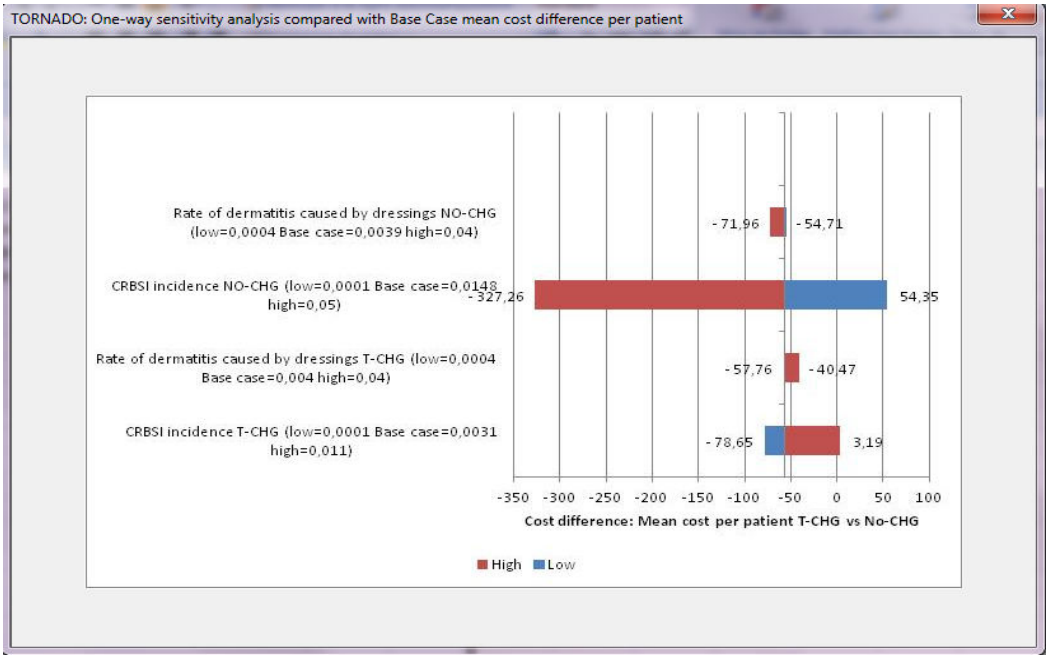

Figure 2. One-way Sensitivity Analysis: Mean cost difference per global patient between CHG and No-CHG strategies, in euros; part 2.

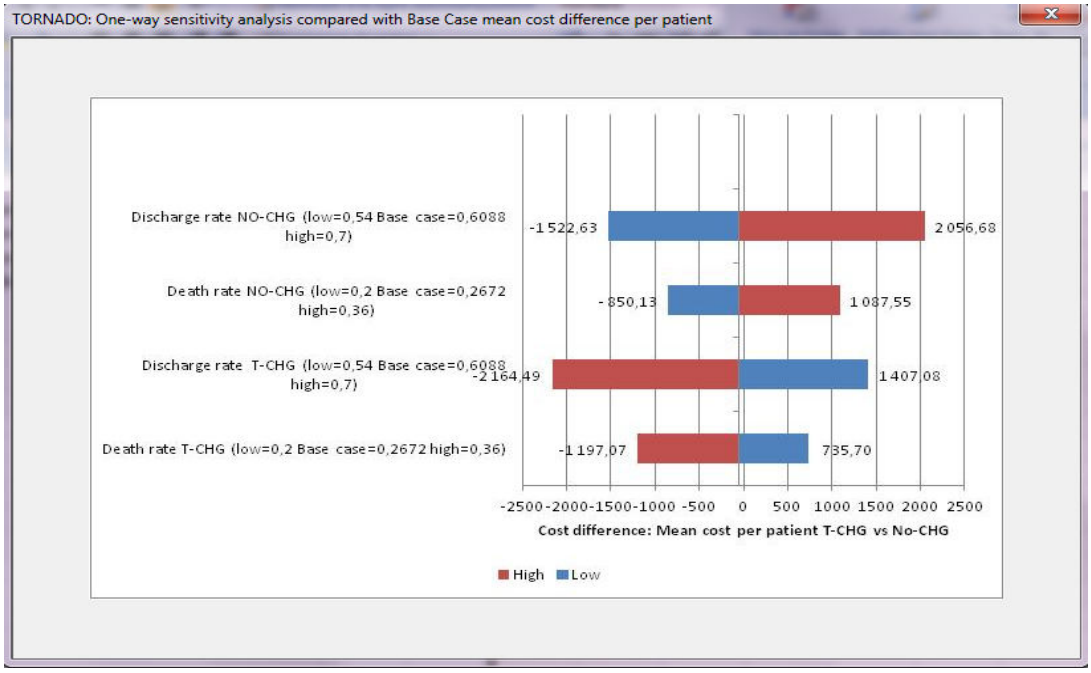

Figure 3. One-way Sensitivity Analysis: Mean cost difference per global patient between CHG and No-CHG strategies, in euros; part 3.

Probabilistic Sensitivity Analysis (PSA)

A probabilistic sensitivity analysis [78] was performed with 1,000 homogeneous MCMC simulations of 1,000 patients for both CHG dressing and non-CHG dressing groups. Each group of 1,000 patients depicts an average patient representing all patients for each dressing group studied in the RCT [44]. The method used was the Gibbs sampling [79], a commonly used Markov Chain Monte Carlo algorithm. It allowed to retrace 10^6 health trajectories (1000*1000 patients for each dressing strategy), based on the probabilities observed in the RCT [44] day-after-day (during 30 days) for each patient to change from one health-state to another. Repeating the algorithm 1,000 times allows the calculation of 95% confidence intervals for the cost-effectiveness criterion (here, number of CRBSI avoided and cost per patient).

The health states including CRBSI (CRBSI/No new catheter and CRBSI/new catheter) as rare events for both strategies are in the area of low probabilities. On the other hand, the “discharge” and “death” states as frequent events for both strategies are in the area

of high probabilities. This corresponds to the reality observed in the RCT (higher frequency of discharge and death than CRBSI).

The CHG-dressing prevents 6.5 infections / 1,000 patients (95% CI: [-12.57; -0.43]) as estimated via probabilistic cost-effectiveness sensitivity analysis in the proposed CEA. The mean adjusted cost per "global" patient is € 21,770 (95% CI: [€20,925; €22,616]) for the CHG-dressing group and €21,819(95% CI: [€20,935; €22,704]) for the reference dressing. Mean cost difference per "global" patient is €-49: (95% CI: [€-1,252; €1,153]) (See Table 7 and Table 8).

Table 7. Mean Number of CR-BSI for 1,000 patients in each dressing group – Horizon: 30-days ICU – 1,000 MCMC simulations of 1,000 patients

Stat.	CHG group (1)	No-CHG group (2)	Diff. Effectiveness (1-2)
Mean	1.8	8.30	-6.50
Lower 95%CI	0	2.77	-12.57
Upper 95%CI	4.44	13.83	-0.43

We can see that the difference of CRBSI events between the strategies is statistically significant at the 0.05-level.

Table 8. Mean Cost for 1 patient in each dressing group – Horizon: 30-days ICU – 1,000 MCMC simulations of 1,000 patients

Stat.	CHG group (1)	No-CHG group (2)	Diff. Cost (1-2)
Mean	€21,770	€21,819	€-49
Lower 95%CI	€20,925	€20,935	€-1,252
Upper 95%CI	€22,616	€22,704	€1,153

So the difference of costs between the strategies is not statistically significant at the 0.05-level.

The PSA cost-effectiveness plan (Figure 4) describes the effectiveness difference on the x-axis and the cost difference on the y-axis between the two groups of dressings, for 1,000 H-MCMC simulations of 1,000 patients in each group. The (0,0)-point indicates the reference dressing strategy (Non-CHG group). All the points observed on the graph represent the incremental cost-effectiveness ratio (ICER) of CHG-dressing strategy versus reference dressing. This PSA supports the decision to adopt

CHG dressing for critically ill patients since the strategy is 97.95% more effective than the comparator at the same cost per patient in the intensive care unit. The mean ICER calculated from the PSA and defined as the cost per patient treated with chlorhexidine dressing to prevent one patient experiencing a CRBSI, is €12,094.

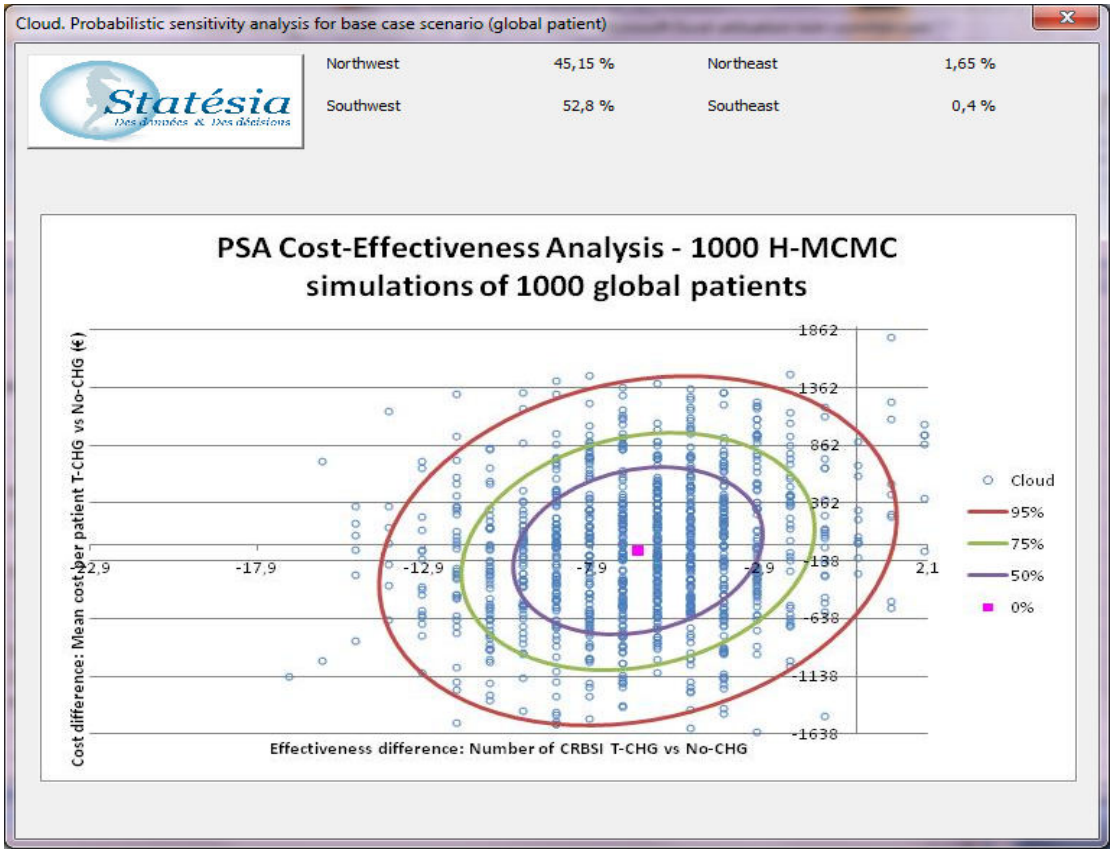

Figure 4. Cost-effectiveness results for the probabilistic sensitivity analysis using 1,000 homogeneous Markov-Chain Monte Carlo simulations of 1,000 patients.

3.4.2 Markov non-homogeneous model

In non-homogeneous case the Markov property is still unchanged. We can state that

$$\mathbf{P}(X_{n+1} = j \mid X_1 = i_1 , X_2 = i_2 ,\dots , X_n = i_n) = \mathbf{P}(X_{n+1} = j \mid X_n = i_n)$$

depends only on i_n, j and n. It means that the probability of the transition depends on n (time) or the probability of the transition changes after each step.

3.4.2.1 Study design

The adopted modeling approach complies with the guidelines of French National Authority for Health (Haute Autorité de Santé - HAS) [67]. The 30-day ICU-time non-homogeneous Markov model [68], [69] structure was based on observed data of a multicentre RCT [44], conducted by the Grenoble University Hospital - CHU Grenoble. The model has been programmed using Visual Basic Application with the Excel® 2007 software. This model consists of eight health states described in Table 9: four states combining either occurrence, or no occurrence of CRBSI, and the need, or no need, of a new central line (CT); one state for contact dermatitis; one for changing to an alternative dressing (gauze and tape) in case of dermatitis, and two absorbing states (death and discharge of the ICU).

Table 9. Health states for NH-MCMC model defined from a multicentre randomized controlled trial

Health States	Definition
1. No CRBSI / No new CT needed	Insertion of a first catheter, no diagnosed CRBSI and no contact dermatitis
2. No CRBSI / new CT needed*	No diagnosed CRBSI, no contact dermatitis and a new catheter inserted (not as a replacement)
3. CRBSI / No new CT needed	CRBSI diagnosed without neither contact dermatitis nor the need for inserting a new catheter
4. CRBSIs / new CT needed*	CRBSI diagnosed without contact dermatitis but the need for inserting a new catheter
5. Contact dermatitis	No diagnosed CRBSI, and no need for new catheter inserted but occurrence of contact dermatitis
6. Dressing Gauze and Tape	Change to an alternative dressing strategy (gauze and tape)
7. Discharge	Patient leaves the ICU alive
8. Death	Patient dies during the ICU stay

* New CT needed can mean either the replacement of the existing catheter, or the need for an additional catheter at a new site.
CRBSI: Catheter-related Bloodstream Infections; CT Catheter: Central venous or radial / femoral arterial

The statistical unit of the study is the ICU patient within a time horizon of 30 days (discharged alive from the ICU, alive but still at the ICU, or deceased during the ICU stay). Patient data from the multicentre RCT [44], comparing the CHG dressing to

non-antimicrobial transparent dressings, were translated into a daily patient transition matrix among the different possible health states, for both the antimicrobial and non-antimicrobial dressings groups. The transition matrixes were used to perform non-homogeneous Markov-Chain Monte Carlo (NH-MCMC) simulations [70] representing the observed daily evolution of patients in ICU. 1,000 Monte Carlo simulations of 1,000 patients were used for probabilistic sensitivity analysis and 95% confidence intervals (CI) calculations.

The final health outcome of the cost-effectiveness analysis is the number of CRBSIs avoided and the cost-effectiveness criterion is the cost per patient with CRBSI avoided resulting from chlorhexidine dressing use.

3.4.2.2 Model structure

Markov models consider patients in a discrete state of health [71], [72], [73], and events representing the transition from one health state to another. This type of model allows an accurate representation of the evaluated clinical structure by modeling repetitive events and time dependence of probabilities (time-nonhomogeneous Markov Chains). The Markov property refers to the fact that the conditional probability distribution of future states of the health depends only upon the present state, not on the sequence of events that preceded it.

In a non-homogeneous modeling approach, the transition probability from one state to the next will change with time, as observed in real life of patients in ICU. The probability of changing from one state to the other can be assembled into a transition matrix.

Building up the transition matrix
A transition matrix for each day in the ICU was built based on transition probabilities reported in the RCT database [44]. Each patient will be in one of the eight health states described in Table 9 at each day in ICU.

The possible transitions among health states from one day to the next are represented in the Markov diagram (see figure below) was censored beyond 30 days. The current model comprises a time horizon of 30 days in ICU, each day corresponding to 1 Markov cycle.

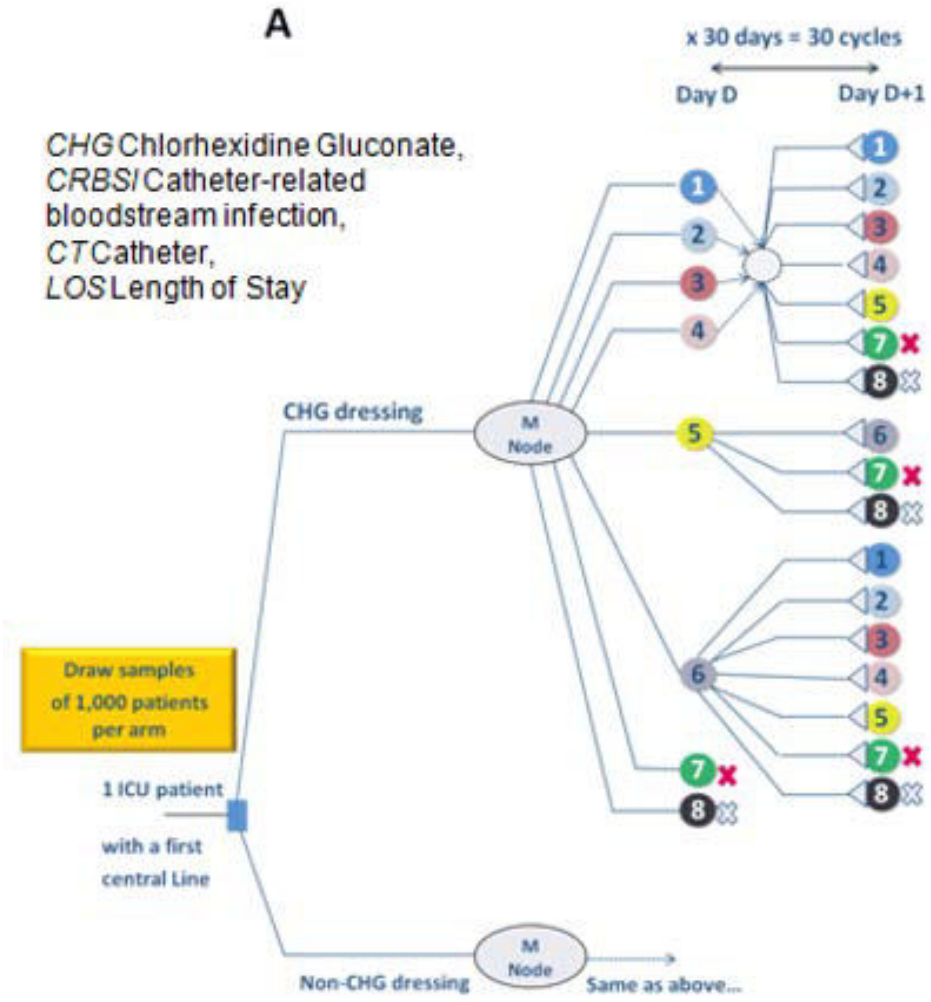

Figure 5. The structure of the Markov Model showing the possible transition between health states from one Markov cycle to the next cycle.

The model structure (see Figure 5) is somewhat based on that published by the independent assessments for managing the CR-BSI. However, more detailed levels are added looking at the "day to day" individual data during the ICU stay. In the Markov model there is no CR-BSI related death state as no evidence was found to support this direct elevated risk of dying.

Unlike the published model, an attempt has been to segregate the rather heterogeneous "day to day" patient health state in CR-BSI control.

3.4.2.3 Healthcare resource use and costs

Costs used within the model reflect the ICU perspective in France and consist of following components:

- drug acquisition costs,
- cost of treating the adverse events ,
- direct costs of treating the CR-BSI,
- costs due to ICU stay,
- cost per catheter change.

Base case input parameters considered in the cost analysis

The base case analysis is the most representative case of the real life, considering French ICU settings, and depending on expert opinions, literature, and RCTs.
The main input parameters considered in the cost analysis are the following:

- Dressing costs per day: CHG dressing which is 20 times more expensive than non-antimicrobial transparent film and 60 times costly than gauze and tape.
- Cost of treating contact dermatitis (mean/episode): catheter removal, 24€ [75]; four gauze and tape dressings, 0.24 €; catheter insertion, 94.87 €.
- Direct cost of treating CRBSI (mean/episode)[75]: 580.26 €.
- Cost per ICU [76]: 1,265.93 €/day.
- Additional ICU Length of stay (LOS) due to CRBSI: 9.33 days (NH-MCMC calculation).
- Cost of added ICU LOS due to CRBSI: 11,811.13 € (NH-MCMC calculation).
- Cost per catheter change (venous + arterial: 50/50 %) [75]: 94.97 €
- Overall cost of one CRBSI (direct cost of treating one CRBSI plus cost of additional ICU LOS due to CRBSI): 12,391.40 € (calculation).

Direct costs for the treatment of CRBSIs were obtained from a micro-costing study. ICU costs were based on an observational (real life) study [76] that assessed all resources consumed during a patient day in the ICU. This twenty-four hours multicentre prospective medico-economic study provides a complete overview and estimation of the actual average cost for medical and surgical ICUs in different hospital types in France: Hospitals (CH), University Hospitals (CHU) and Regional Hospitals (CHR). Twenty-two ICUs were selected randomly and all costs for 109 patients were

estimated. For patients with CRBSI, an additional cost [77] due to an extra ICU length of stay (LOS) was calculated (see next section).

Main Assumptions Used for the Cost Analysis

- The cost of CR-BSI is independent from the outcome (survival or death or discharge). For the analysis the main statistical unit is the "global" (survival or death or discharge) patient;
- Catheter colonization with or without CR-BSI had no costs (after the diagnosis) or adverse outcomes (colonization has been initially excluded from the model because it's not a "health-state". The costs for diagnosis are negligible compared to costs related to additional LOS. The costs related to replace a catheter suspected to be colonized (and causing CR-BSI) will be absorbed in the health-states CT new (the dressing arm with NO CHG);
- The estimated cost per ICU day at the Grenoble University Hospital is identical in each Dressing Group;
- The G+T (gauze and tape) cost for CHG-group is identical to No-CHG group.

Costs items for each Markov state

Table 10 reportes the costs included in the analysis for each Markov state.

Table 10. Costs items for each Markov state

Main costs	**Detailed costs / Data Provider**	no CR-BSI/ no CT new	no CR-BSI/ CT new	CR-BSI/ no CT new	CR-BSI/ CT new	Contact dermatitis	Death	Discharge
Cost of dressings	*CHU Grenoble*	X	X	X	X	X		
	Time needed per dressing	X	X	X	X	X		
	Number of nurses involved	X	X	X	X	X		
	Material used	X	X	X	X	X		
Cost of treating contact dermatitis	*CHU Grenoble/ Dressing 1/ Schwebel 2012*					X		

	Four G+T					X		
	Removal of the catheter					X		
	Insertion of a new catheter					X		
Cost of treatment of CR-BSI	*CHU Grenoble/ Dressing 1/ Schwebel 2012*			X	X			
Additional ICU -LOS due to CR-BSI	*Dressing 2/ Statésia*			X	X			
Cost per ICU day	Garrigues 2010	X	X	X	X	X		
VM (mechanical ventilation)	*CHU Grenoble/Schwebel 2012*	X	X	X	X	X		
Inotrope	*CHU Grenoble/Schwebel 2012*	X	X	X	X	X		
Hemodialysis	*CHU Grenoble/Schwebel 2012*	X	X	X	X	X		
Hemofiltration	*CHU Grenoble/Schwebel 2012*	X	X	X	X	X		
Forfait journalier REA	*CHU Grenoble/Schwebel 2012*	X	X	X	X	X		
Cost per catheter change (Venous, arterial)	Schwebel 2012		X		X			

Costs per Markov state per patient

The calculation of the cost for each Markov state per patient was done as follows (using the base case input parameters listed above):

- Dressing costs (including time needed per dressing, number of nurses involved, and materials used [75]) and cost per ICU day [76] were taken into account for health states 1-6;
- Cost of treating contact dermatitis [75] – (including catheter removal, four alternative dressings, and insertion of a new catheter) was taken into account only for health state 5;
- Cost of treatment of CRBSI [75] and additional ICU-LOS due to CRBSI [44], [75] were taken into account for health states 3 and 4;
- Cost per catheter change (venous, arterial) [75] was taken into account for health states 2 and 4.

The costs per patient for each health state were calculated in both CHG and No-CHG dressing and presented in Table 11.

Table 11. Costs per Markov state per patient from the base case scenario

	Costs for 1 patient CHG, Euro 2013	***Costs for 1 patient No-CHG, Euro 2013***
NoAE/noCRBSI/noCTnew	1,270	1,266
NoAE/noCRBSI/CTnew	1,364	1,361
NoAE/CRBSI/noCTnew	13,661	13,658
NoAE/CRBSI/CTnew	13,756	13,752
Contact dermatitis	1,388	1,385
Dressing G+T	1,266	1,266
Discharge	0	0
Death	0	0

3.4.2.4 Results of cost-effectiveness analysis

The results presented below are from the base case scenario of the cost-effectiveness modeling. The structure follows the next principal sections:

- Base case scenario results for 1,000 NH-MCMC of 1,000 patients CHG group
- Base case scenario results for 1,000 NH-MCMC of 1,000 patients No-CHG group

- Number of events for each state for 1,000 NH-MCMC of 1,000 patients CHG group – Time Horizon: 30-days ICU
- Number of events for each state for 1,000 NH-MCMC of 1,000 patients No-CHG group – Time Horizon: 30-days ICU

The results showed in Tables 12-16 refer to the base case scenario of the cost-effectiveness modeling. Tables 12 and 13 show that the main difference between CHG group and Non-CHG group are based on % of states 3 and 4, e.g. CRBSIs. A ratio of 1 to 5 is observed for the average number of CRBSIs between dressing groups. The number of ICU-days, the number of days before discharging and the number of days before dying are comparable in the two groups.

Table 12. Base case scenario results for 1,000 NH-MCMC of 1,000 patients CHG group

Statistics	Mean	Lower 95%CI	Upper 95%CI
State 2 No AE/no CRBSI/CT new	27.82%	24.18%	31.45%
State 3 No AE/CRBSIs/no CT new	0.00%	0.00%	0.00%
State 4 No AE/CRBSIs/CT new	0.31%	0.00%	6.48%
State 5 AE/no CRBSI/no CT new	2.88%	1.46%	4.30%
Number of ICU-days	12.91	12.30	13.52
Number of days before State 7 Discharge	18.74	18.05	19.43
Number of days before State 8 Death	25.17	24.49	25.85

Table 13. Base case scenario results for 1,000 NH-MCMC of 1,000 patients Non-CHG group

Statistics	Mean	Lower 95%CI	Upper 95%CI
State 2 No AE/no CRBSI/CT new	25.16%	21.88%	28.44%
State 3 No AE/CRBSIs/no CT new	0.53%	0.07%	0.98%
State 4 No AE/CRBSIs/CT new	0.95%	0.33%	1.57%
State 5 AE/no CRBSI/no CT new	1.27%	0.44%	2.09%
Number of ICU-days	12.72	12.12	13.32
Number of days before State 7 Discharge	18.43	17.72	19.16
Number of days before State 8 Death	25.28	24.64	25.92

CRBSI occurred for 3 and 14 patients in each CHG and non-CHG groups respectively (1,000 patients in each group; Tables 12 and 13). This difference was highly statistically significant as indicated by the non-overlapping 95% confidence intervals (see Figure 6).

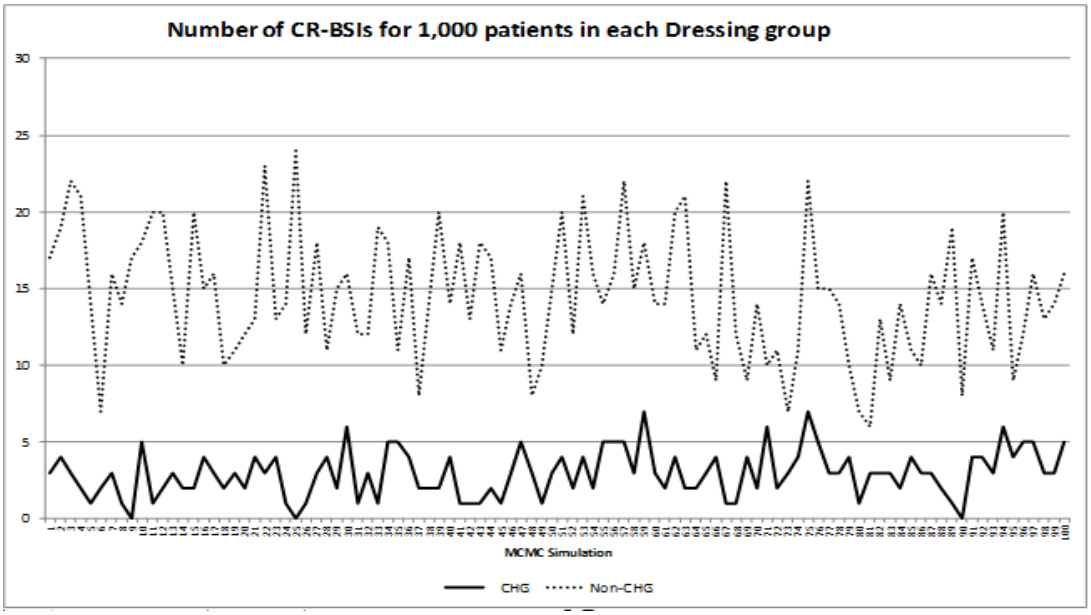

Figure 6. Number of CRBSIs for 1,000 patients in each dressing group (results for the first 100 of the 1,000 simulations).

The percentage of patients in the two absorbing states, coded as 7 (discharge from ICU) and 8 (death), is comparable in both groups of dressings, (see Table 14, Table 15 and Figure 7).

Table 14. Percentage of events for absorbing states for 1,000 NH-MCMC of 1,000 patients CHG group – Time Horizon: 30-days ICU

Statistics	% of State 7 (ICU Discharge)	% of State 8 (Death)
Mean	60.41%	26.37%
Lower 95%CI	57.44%	23.47%
Upper 95%CI	63.38%	29.27%

Table 15. Percentage of events for absorbing states for 1,000 NH-MCMC of 1,000 patients Non-CHG group – Time Horizon: 30-days ICU

Statistics	% of State 7 (ICU Discharge)	% of State 8 (Death)
Mean	61.34%	27.07%
Lower 95%CI	58.28%	24.24%
Upper 95%CI	64.41%	29.90%

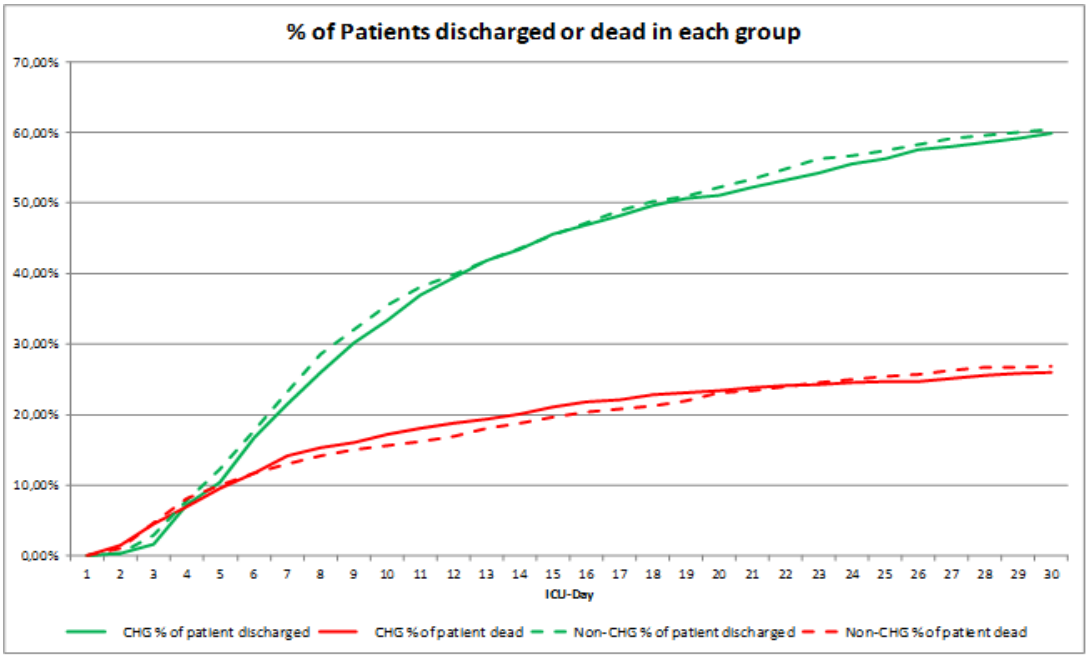

Figure 7. Percentage of discharged or dead patients in each dressing group.

Considering a defined global average patient cohort, the average cost per patient per ICU-day decreases with time due to the increasing percentage of discharged or dead patients during the ICU stay (the cost for a discharged or dead patient is considered as zero in the model). The curves are showing the average cost considering an initial population of defined sized (Figure 8).

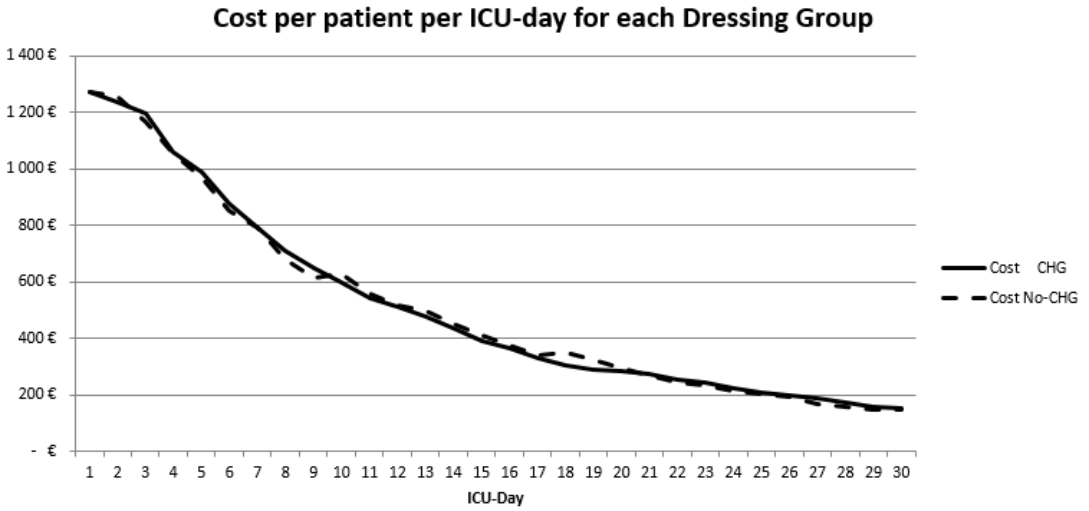

Figure 8. Cost per patient per ICU-day for each dressing group.

Table 16 shows the cost results for the average patient in each dressing group.

Table 16. Mean Cost for one patient in each dressing group. Time Horizon: 30-days ICU – 1,000 NH-MCMC simulations of 1,000 patients

Groups /Statistics	Mean	Lower 95%CI	Upper 95%CI
ALL PATIENTS			
CHG (1)	€16,461	€15,659	€17,265
Non-CHG (2)	€16,320	€15,538	€17,103
Diff. Cost (1-2)	€141	€-975	€1,258
PATIENTS with CRBSI in ICU			
CHG (1)	€39,071	€17,384	€60,758
Non-CHG (2)	€41,424	€36,213	€46,635
Diff. Cost (1-2)	€-2,353	€-24,984	€20,277
PATIENTS without CRBSI			
CHG (1)	€16,385	€15,584	€17,186
Non-CHG (2)	€15,946	€15,177	€16,715
Diff. Cost (1-2)	€439	€-664	€1,542

For a 30-day time horizon in ICU, the mean cost per patient for CHG group was of €16,461, versus €16,320 for the non-CHG strategy. The mean cost per patient with CRBSI was of €39,071 and €41,424 in CHG and non-CHG dressing groups while the mean cost per patient without CRBSI was of €16,385 and €15,946 in CHG and non-CHG dressing groups, respectively (see Table 16). Subgroup analyses supported by the comparability test compared the average total costs for patients with CRBSI versus patients without CRBSI for each study group (CHG and Non-CHG dressings). This comparison revealed no significant differences in costs among the subgroups (Table 16). Figure 9 shows the non-overlapping 95% confidence intervals for each dressing strategy.

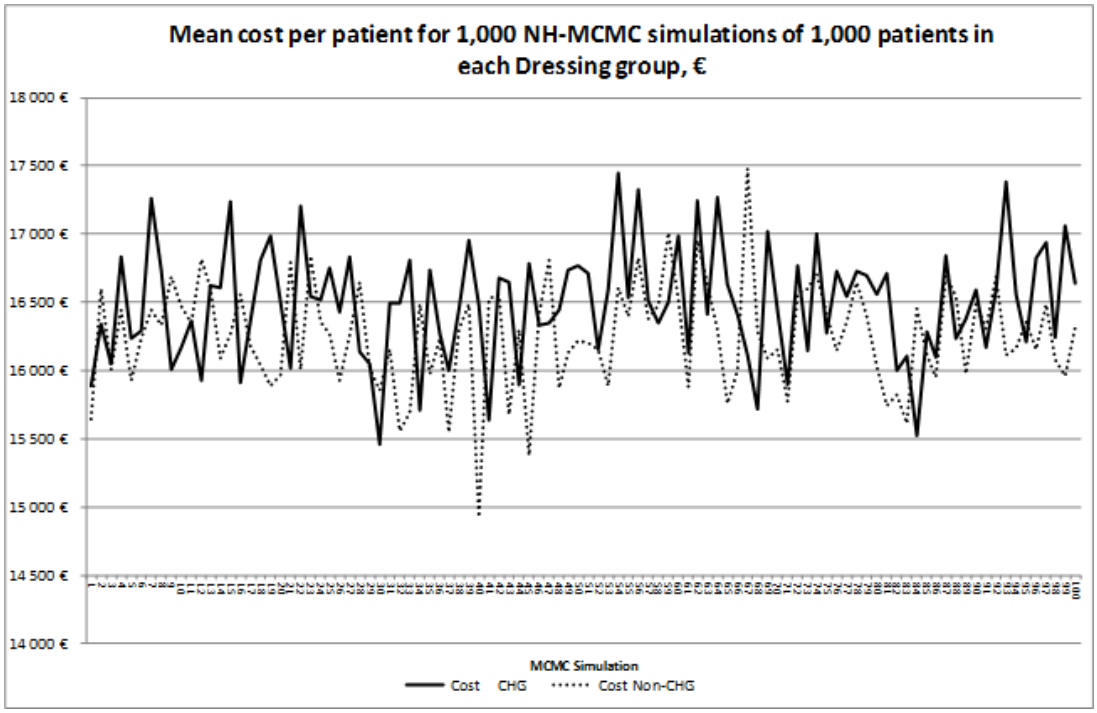

Figure 9. Mean cost per patient in each dressing group (results for the first 100 of the 1,000 simulations).

3.4.2.5 Sensitivity analyses

Sensitivity analyses are performed to vary each parameter of the model in order to determine what levels will result in a change of preference for the therapeutic strategy. This is a way to test the boundaries of the model and identify the main parameters driving cost differences.

One-way Sensitivity Analysis

One-way sensitivity analyses were performed varying the main input parameters (additional ICU LOS due to CRBSI (days), CHG Dressing cost, number of CHG dressing per day, number of Non-CHG dressing per day, and cost per ICU day) of the model around the base case assumptions.

A tornado diagram (Figure 10) shows the variation in the mean cost difference between the CHG and non-CHG strategies around the one calculated for the base-case (€141). The model was most sensitive to the variation of the number of extra ICU LOS due to CRBSIs. The cost difference varied of approximately €370, when accounting

from a single extra ICU day (cost difference of €251) to 26 extra ICU days (cost difference of €-115). The next three influential parameters were the CHG-dressing cost, the interval for dressing change, and the cost per ICU-day. However, the variation in the cost differences obtained by changing these parameters was less pronounced (differences between upper and lower limits of 88, 85 and €83, respectively).

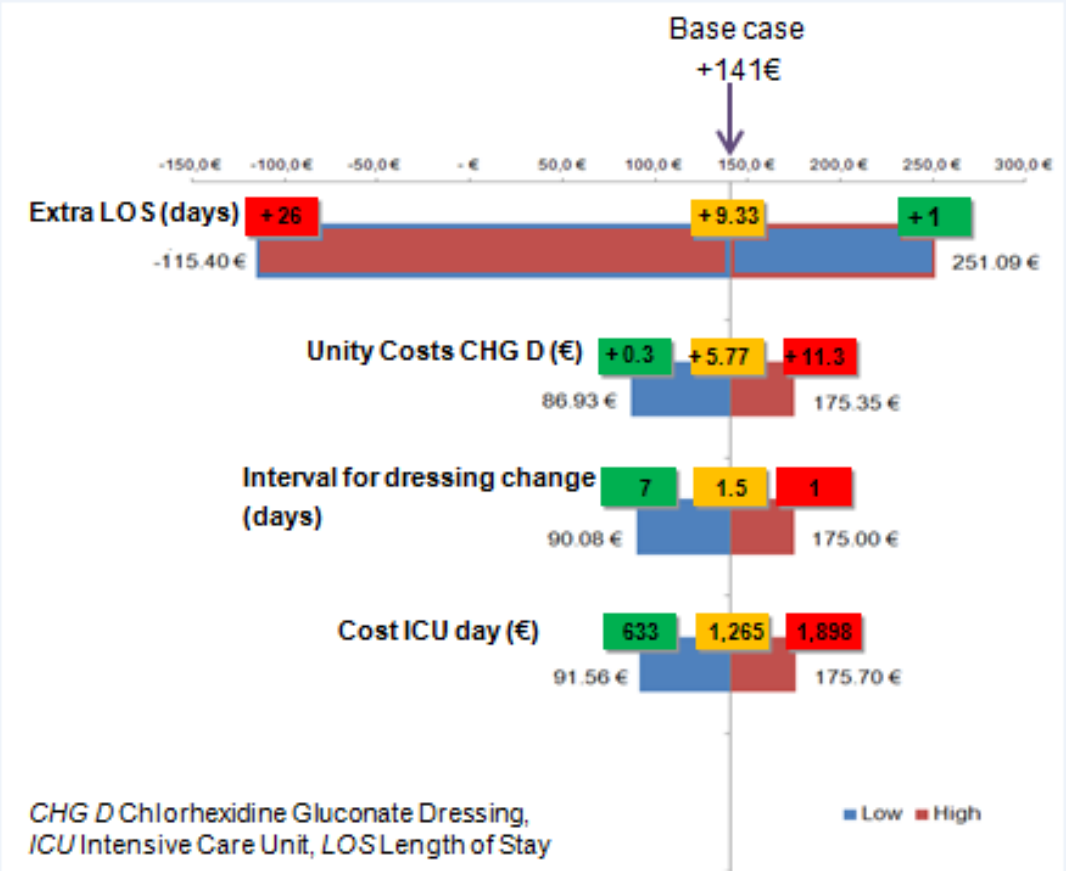

Figure 10. Tornado diagram for the one-way sensitivity analysis.

Probabilistic Sensitivity Analysis (PSA)

A probabilistic sensitivity analysis [78] was performed with 1,000 non-homogeneous MCMC simulations of 1,000 patients for both CHG dressing and non-CHG dressing groups. Each group of 1,000 patients depicts an average patient representing all patients for each dressing group studied in the RCT [44]. The method used was the

Gibbs sampling [79], a commonly used Markov Chain Monte Carlo algorithm. It allowed to retrace 10^6 health trajectories (1000*1000 patients for each dressing strategy), based on the probabilities observed in the RCT [44] day-after-day (during 30 days) for each patient to change from one health-state to another. Repeating the algorithm 1,000 times allows the calculation of 95% confidence intervals for the cost-effectiveness criterion (here, number of CRBSI avoided and cost per patient).

The health states including CRBSI (CRBSI/No new catheter and CRBSI/new catheter) as rare events for both strategies are in the area of low probabilities. On the other hand, the "discharge" and "death" states as frequent events for both strategies are in the area of high probabilities. This corresponds to the reality observed in the RCT (higher frequency of discharge and death than CRBSI).

The PSA cost-effectiveness plan (Figure 11) describes the effectiveness difference on the x-axis and the cost difference on the y-axis between the two groups of dressings, for 1,000 NH-MCMC simulations of 1,000 patients in each group. The (0,0)-point indicates the reference dressing strategy (Non-CHG group). All the points observed on the graph represent the incremental cost-effectiveness ratio (ICER) of CHG-dressing strategy versus reference dressing. This PSA supports the decision to adopt CHG dressing for critically ill patients since the strategy is 99.8% more effective than the comparator at the same cost per patient in the intensive care unit. The mean ICER calculated from the PSA and defined as the cost per patient treated with chlorhexidine dressing to prevent one patient experiencing a CRBSI, is €12,046.

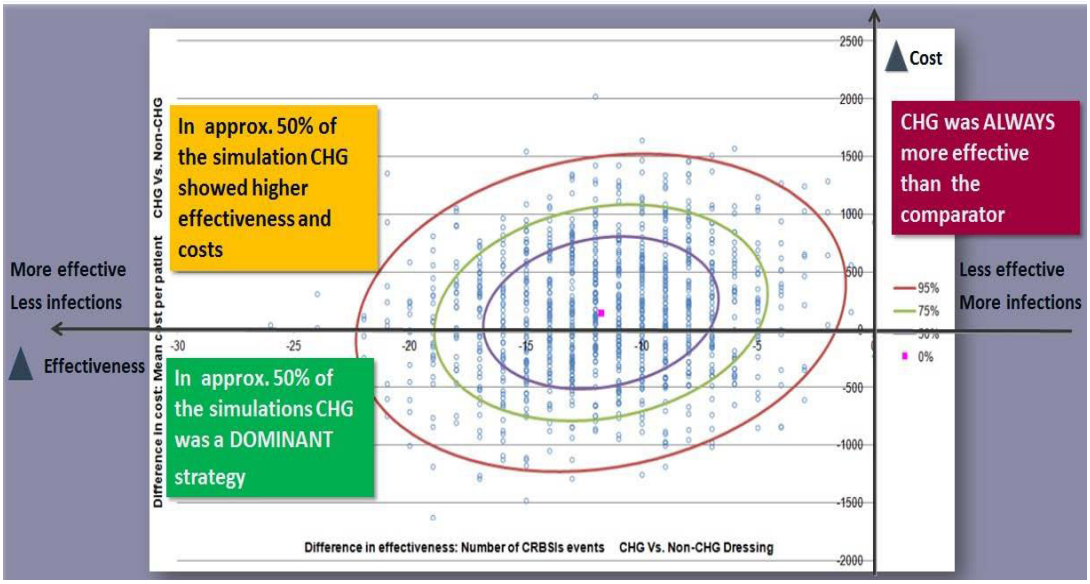

Figure 11. Cost-effectiveness results for the probabilistic sensitivity analysis using 1,000 non-homogeneous Markov-Chain Monte Carlo simulations of 1,000 patients.

Convergence of MCMC

A large number of methods of convergence diagnostic methods for Markov Chain Monte Carlo algorithm (MCMC) is available in the literature [80], [81], [82].

To illustrate the convergence of MCMC consider that we will estimate only the cumulative number of patients that develop a CR-BSI during 30 days in the ICU using CHG dressing. This event correspond to the rarest event represented in our study. Note that the true value of this parameter is 3 patients for 934. The convergence of all other events in our work could be treated in the same way.

In order to reduce the possibility of bias due to the effect of starting values for Markov Chains and "burn in" period we conducted a Monte Carlo simulations of 1,000 replicates, each with N=1,000 individuals (1,000,000 individuals overall).

Here we cannot perform just one long run of single chain as the convergence times are not available [83], [84]. The option is to take several shorter runs of a number of independent chains and form a sample from these observations [85]. In the database the number of patients in CHG and No-CHG group was 934 and 935. So the choice of sampler for Markov Chain that is 1,000 patients is coherent. The number of independent replications to be run should be estimated.

We propose to run several chains in parallel and to estimate the parameter of interest (see Figure 12 and Figure 13). The convergence can be achieved when the estimations of parameter from the individual chains become indistinguishable. We can see that had we only 100 or 500 iterations of our Markov chain we might easily underestimate or overestimate the occurrence of CR-BSI event. Based on 1,000 iterations the estimator of our parameter converge to the true value. By continuing the number of replications beyond 1,000 (in graphs 2,000 and 5,000 iterations) we can be reasonably certain that convergence of our Markov chain was achieved after 1,000 iterations.

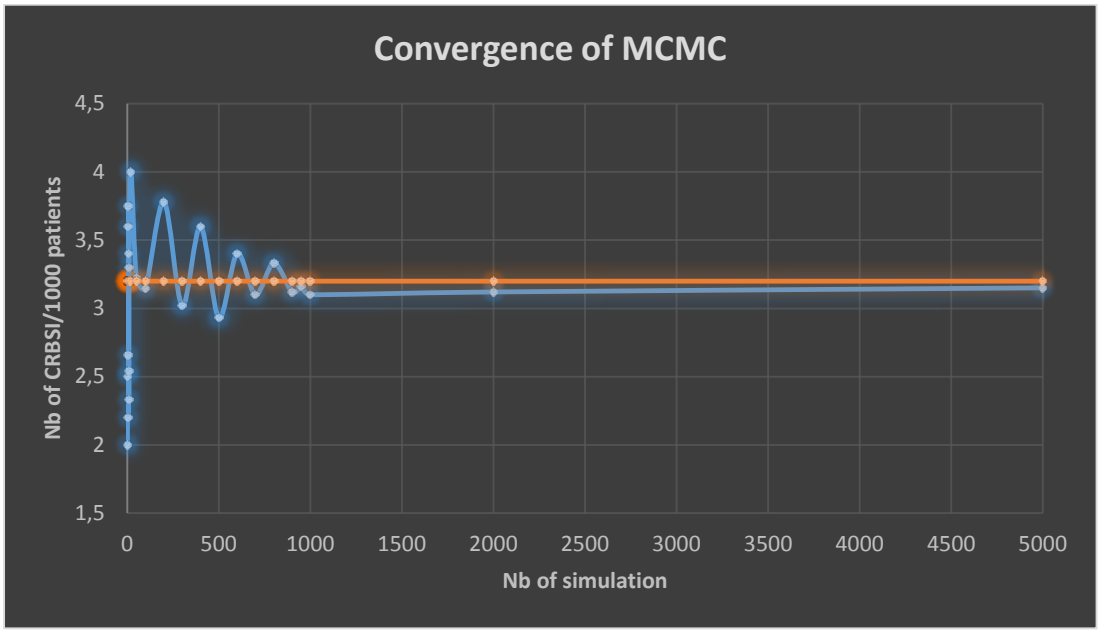

Figure 12. Convergence of Markov Chain

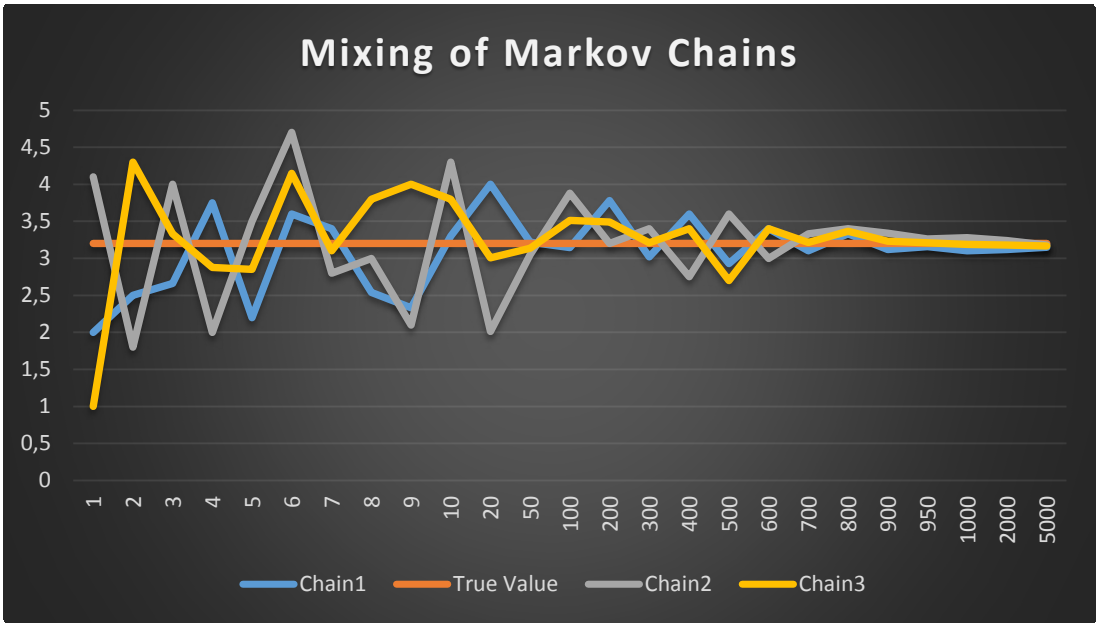

Figure 13. Mixing of Markov Chains

Incremental net monetary benefit

The average incremental net monetary benefit (iNMB) of €344 per patient induced by a daily CHG dressing use in patients resident in intensive care units was calculated, (Table 17).

Table 17. Incremental net monetary benefit (iNMB) for 1 patient induced by CHG use – Horizon: 30-days ICU – 1,000 NH-MCMC simulations of 1,000 patients

Statistics (iNMB)	Cost-effective, if positive sign
Mean	€344.88
Lower 95%CI	€-883.01
Upper 95%CI	€1,572.77

The iNMB is calculated from the difference between the cost induced by CHG dressing use (mean cost CHG patient – mean cost Non-CHG patient) and the cost averted by CHG use (differential of effectiveness per patient x mean cost per CRBSI Non-CHG patient). If mean cost per CRBSI Non-CHG patient is considered as the current willingness to pay (WTP) for treating one patient with CRBSI, the incremental net monetary benefit of CHG dessing use for the ICU is as follow:

$$iNMB = \Delta\, Effectiveness * WTP - \Delta\, Cost.$$

If the “incremental net monetary benefit” is positive, this indicates that the assessed technology is cost-effective.

3.4.3 Comparison of two models. Discussion

In this work, we performed the two approaches homogeneous and non-homogeneous Markov models.

The first homogeneous (time-independent) approach which is closely associated with the non-homogeneous Markov model developed after. In each dressing strategy (CHG, No-CHG), homogeneous MCMC simulations were performed to estimate the cost-effectiveness results. This homogeneous Markov chain Monte Carlo (H-MCMC) simulation was carried out with the statistical unit which is described in this study report as “global patient”. The “global patient” is closest to the observed “ICU patient’s health trajectories”, he could be alive, discharged or dead during his ICU stay.

The CHG-dressing prevents 6.5 infections / 1,000 patients (95% CI: [0.43; 12.57]) as estimated via probabilistic cost-effectiveness sensitivity analysis in the proposed CEA. The mean adjusted cost per “global” patient is €21,770 (95% CI: [€20,925; €22,616]) for the CHG-dressing group and €21,819 (95% CI: [€20,935; €22,704]) for the reference dressing. Mean cost difference per “global” patient is €-49:
(95% CI: [€-1,252; €1,153]).

The second model that we performed is a non-homogeneous Markov model which is very linked to the Dressing 2 individual observed data. This time-dependent transition model allows simulating, with a high goodness of fit, the “health trajectory” of individuals observed in the Dressing 2 study.

The CHG-dressing prevents 11.75 infections /1,000 patients (95% CI: [3.85;19.64], number needed to treat = 85) as estimated via probabilistic cost-effectiveness sensitivity analysis. The mean adjusted cost per “global” patient is € 16,462 (95% CI: [€15,659; €17,265]) for the CHG-dressing group and €16,320 (95% CI: [€15,538; €17,103]) for the reference dressing. Mean cost difference per “global” patient is +€141 (95% CI: [€-975; €1,258] and mean net saving per patient is €-344.88 (95% CI: [€-1,572.77; €883.01]).

Mean net saving is calculated as follow: Cost induced – Cost averted = (Mean cost CHG patient – Mean cost No-CHG patient) – (mean effectiveness difference per patient x mean cost per CR-BSI No-CHG patient).

This non-homogeneous Markov Chain Monte Carlo (NH-MCMC) simulation was carried out with the statistical unit which is described in this study report as "global patient". Therefore, to take into account the "all causes death" state in our modeling decreases the effectiveness of the assessed new CHG-dressing because of the death rate is very important in ICU setting. Indeed, the NH-MCMC simulation on the "global patient" estimated 7 No-CHG died patients and 9 No-CHG discharged patients more than in the CHG group. As these two states are the "absorbent" states of the model, their cost are zero. As a consequence, our results could be considered as a "conservative" scenario.

For taking into account this matter, we performed a NH-MCMC simulation subgroup analysis with alive patients only. The effectiveness result is -9.44 /1,000 patients (95% CI: [-16.58; -2.29]) CR-BSI averted; cost difference per alive patient is +€618.43 (95% CI: [€-725.16; €1,962.01]) and mean net over cost per alive patient is +€226.39 (95% CI: [€-1,215.32; €1,668.11]), comparatively to the No-CHG group. This result is based on the fact that mortality is higher in the No-CHG group (see the results above) and this outcome is considered as zero-cost absorbing state. Therefore, we could achieve another NH-MCMC simulation subgroup analysis with patients neither dead nor discharged, but, in this sample, no CRBSI event was recorded.

In each Dressing Strategy (CHG, No-CHG), a non-homogeneous MCMC simulation was performed for a subgroup analysis: "Global" patients with CR-BSI versus "Global" patients without CR-BSI. These NH-MCMC simulations were based on observed patient data, collected during the Dressing 2 clinical study. In this study, all patients were randomly assigned to one of the two dressing strategies, what allow us to assume comparability of the groups.

Discussing further on the comparability of these two subgroups, we can present two additional adjustments.

First, a "natural" adjustment, linked to the main statistical unit of our modeling which is the "global" patient was considered. The probability of developing a CR-BSI in the

"Global population", that corresponding to the clinical trial population, follows the same plausible statistical distribution law (due to the randomization).

The statistical analysis for all confounding covariates, such as age, sex, severity (SOFA score), duration of catheterization, number of dressing change per day, shows the comparability between these subgroups. Nevertheless, a potential limit of the model can come from the covariates which are not available in the database. Such variables as smoking status, obesity, diabete, etc. could induce the important bias.

A further adjustment from NH-MCMC simulation was performed, considering covariates which could impact mainly the values closely linked to the cost-effectiveness results. The rate of catheter change and the number of additional ICU days due to CRBSI in each dressing strategy were taken into account.

Comparison of two approaches

The difference in clinical outcomes between each dressing strategies was statistically significant with both models while cost differences were not. The PSA with the NH-MCMC resulted in 11.8 infections avoided per 1,000 patients (95%CI:[3.85; 19.64]) and a mean extra cost of €141 per patient (95%CI: [€-975; €1,258]) when using antimicrobial dressing. The PSA with the H-MCMC resulted in 6.45 infections avoided per 1,000 patients (95%CI: [0.15; 12.75]) and the mean extra cost of €252 per patient (95%CI:[€-924; €1,428]).

Comparing the results of two models the effectiveness result of CHG is lower than in No-CHG dressings in the first model. In the H-MCMC model the costs difference is in favor of CHG dressing because the base case scenario eliminates all the differences between the strategies that correspond to the death and discharge health states for ICU patients. In the non-homogeneous model a "conservative" scenario was considered. The NH-MCMC simulation on the "global patient" estimated 7 No-CHG died patients and 9 No-CHG discharged patients more than in the CHG group.

Nevertheless it is still possible to enter the different rates as the inputs in the homogeneous model. Thus, the principal limit of this study is the time-independence of the Markov process that means that we have accepted the same transition probabilities over the time spending in ICU.

According to the base case scenario, the CHG-dressing significantly more efficacious to prevent CRBSI when compared to the reference dressing for the ICU patients, contributes to preserve patients' health capital at the same cost for the ICU.

The antimicrobial dressings are consistently more efficacious in preventing CRBSIs whatever the model used. The H-MCMC is less sensitive to simulate the real life of the ICU patients. Regardless the model approach chosen the antimicrobial strategy is more efficacious than the comparator, but its probability of being cost-effective is comparatively reduced with the H-MCMC. Time dependent approach (NH-MCMC) seems to be better adapted to model rare events as CRBSIs.

Bibliography

1. Billingsley P. Statistical Inference for Markov Processes. Chicago: The University of Chicago Press, 1961, 75 p.

2. Bosq D. Mathematical Statistics and Stochastic Processes. London: ISTE, 2012, 304 p.

3. Ibragimov I.A., Has'minskii R. Statistical Estimation - Asymptotic Theory. New York: Springer-Verlag, 1981, 403 p.

4. Fan J., Yao Q. Nonlinear Time Series: Nonparametric and Parametric Methods. New York: Springer-Verlag, 2003, 576 p.

5. Kamatani K., Uchida M. Hybrid multi-step estimators for stochastic differential equations based on sampled data. To appear in Stat. Inference Stoch. Process., 2015, 18, 2.

6. Kushnir A.F. Asymptotically optimal tests for a regression problem of testing hypotheses. Theory Probab. Appl., 1968, 13, 4, p. 647-666.

7. Kutoyants Y.A. On approximation of the backward stochastic differential equation. Small noise, large samples and high frequency cases. Proc. Steklov Inst. Math., 2014, 287, p. 133-154.

8. Kutoyants Y.A. On multi-step MLE-processes for ergodic diffusion. 2015 submitted.

9. Kutoyants Y.A., Motrunich A. On multi-step MLE-process for Markov sequences. 2015 submitted.

10. Kutoyants Y.A., Zhou L. On approximation of the backward stochastic differential equation. J. Stat. Plann. Infer., 2014, 150, p. 111-123.

11. Le Cam L. On the asymptotic theory of estimation and testing hypotheses. Proc. 3rd Berkeley Symposium, 1956, 1, p. 355-368.

12. Motrunich A. On parameter estimation for Markov sequences. Ann. I.S.U.P, 2015 submitted.

13. Ogata Y., Inagaki N. The weak convergence of the likelihood ratio random fields for Markov observations. Ann. Inst. Stat. Math., 1977, 29, Part A, p. 165-187.

14. Roussas G.G. Asymptotic inference in Markov processes. Ann. Math. Statist., 1965, 36, 3, p. 978-992.

15. Roussas G.G. Contiguity of Probability Measures: Some Applications in Statistics. London: Cambridge University Press, 1972, 264 p.

16. Taniguchi M., Kakizawa Y. Asymptotic Theory of Statistical Inference for Time Series. New York: Springer-Verlag, 2000, 662 p.

17. Varakin A.B., Veretenikov A.Yu. On parameter estimation for ”polynomial ergodic” Markov chains with polynomial growth loss functions. Markov Process. Relat., 2002, 8, 1, p. 127-144.

18. Veretenikov A.Yu. Parametric and Non-Parametric Estimation of Markov Chains [in Russian]. Moscow: Moscow University Press, 2000, 95 p.

19. Maki D.G., Kluger D.M., Crnich C.J. The risk of bloodstream infection in adults with different intravascular devices: a systematic review of 200 published prospective studies. Mayo Clin. Proc., 2006, 81, 9, p. 1159-1171.

20. Raad I. Intravascular-catheter-related infections. Lancet, 1998, 351, 9106, p. 893-898

21. Crnich C.J., Maki D.G. The promise of novel technology for the prevention of intravascular device-related bloodstream infection. Pathogenesis and short-term devices. Clin. Infect. Dis., 2002, 34, 9, p. 1232-1242.

22. Byers K.E., Adal K.A., Anglim A.M., et al. Case fatality rate for catheter-related bloodstream infections: a meta-analysis. Infect. Control Hosp. Epidemiol., 1995, 16, 2, p. 23.

23. Kluger D.M., Maki D.G. The relative risk of intravascular device related bloodstream infections in adults. Presented at the 39th Interscience Conference on Antimicrobial Agents and Chemotherapy, September 26-29, 1999, San Francisco, CA.

24. Pittet D., Tarara D., Wenzel R.P. Nosocomial bloodstream infection in critically ill patients: Excess length of stay, extra costs, and attributable mortality. JAMA-J. Am. Med. Assoc., 1994, 271, 20, p. 1598-1601.

25. Arnow P, Quimonsing E, Beach M. Consequences of intravascular catheter sepsis. Clin. Infect. Dis., 2000, 16, p. 778-784.

26. Mermel L. Prevention of intravascular catheter-related infections. Ann. Intern. Med., 2000, 132, p. 391-402.

27. Mermel L.A., McCormick R.D., Springman S.R. et al. The pathogenesis and epidemiology of catheter-related infection with pulmonary artery swanganz catheters: A prospective study utilizing molecular subtyping. American Journal of Medicine, 1991, 91, 3, p. 197-205.

28. Maki D.G. Infections associated with intravascular lines. Curr. Clin. Top. Infect. Dis., 1996, 3, p. 309-363

29. Wenzel R.P., Thompson R.L., Landry S.M., et al. Hospital-acquired infections in intensive care unit patients: An overview with emphasis on epidemics. Infect. Control, 1983, 4, 5, p. 371-375.

30. Muñoz P., Bouza E., San Juan R., et al. Clinical-epidemiological characteristics and outcome of patients with catheter-related bloodstream infections in Europe. Clin. Microbiol. Infec., 2004, 10, 9, p. 843-845.

31. Blot S.I., Depuydt P., Annemans L., et al. Clinical and economic outcomes in critically ill patients with nosocomial catheter-related bloodstream infections. Clin. Infect. Dis., 2005, 41, 11, p. 1591-1598.

32. Bashir M.H., Olson L.K.M., Walters S. Suppression of regrowth of normal skin flora under chlorhexidine gluconate dressings applied to chlorhexidine gluconate-prepped skin. Am. J. Infect. Control, 2012, 40, 4, p. 344-348.

33. Bjornson H.S., Colley R., Bower R.H., et al. Association between microorganism growth at the catheter insertion site and colonization of the catheter in patients receiving total parenteral nutrition. Surg., 1982, 92, 4, p. 720-727.

34. Eggimann P., Harbarth S., Constantin M.N., et al. Impact of a prevention strategy targeted at vascular-access care on incidence of infections acquired in intensive care. Lancet, 2000, 355, p. 1864-1868.

35. National Nosocomial Infection Surveillance (NNIS) System report: data summary from January 1990 - May 1999. Am. J. Infect. Control, 1999, 27, p. 520-532.

36. Coopersmith C.M., Rebmann T.L., Zack J.E., et al. Effect of an education program on decreasing catheter-related bloodstream infections in the surgical intensive care unit. Crit. Care Med., 2002, 30, 1, p. 59-64.

37. Pronovost P., Needham D., Berenholtz S., et al. An intervention to decrease catheter-related bloodstream infections in the ICU. N. Engl. J. Med., 2006, 355, 26, p. 2725-2732.

38. Pronovost P.J., Goeschel C.A., Colantuoni E., et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: Observational study. BMJ, 2010, 340, 7744, p. 462

39. Chaiyakunapruk N., Veenstra D.L., Lipsky B.A., ey al. Chlorhexidine compared with povidone-iodine solution for vascular catheter-site care: A meta-analysis. Ann. Intern. Med., 2002, 136, 11, p. 792-801.

40. Bleasdale S.C., Trick W.E., Gonzalez I.M., et al. Effectiveness of chlorhexidine bathing to reduce catheter-associated bloodstream infections in medical intensive care unit patients. Arch. Intern. Med., 2007, 167, 19, p. 2073-2079.

41. Popovich K.J., Hota B., Hayes R., et al. Effectiveness of routine patient cleansing with chlorhexidine gluconate for infection prevention in the medical intensive care unit. Infect. Control Hosp. Epidemiol., 2009, 30, 10, p. 959-963.

42. Casey A.L., Mermel L.A., Nightingale P., et al. Antimicrobial central venous catheters in adults: A systematic review and meta-analysis. Lancet Infect. Dis., 2008, 8, 12, p. 763-776.

43. Timsit J., Schwebel C., Bouadma L., et al. Chlorhexidine-lmpregnated sponges and less frequent dressing changes for prevention of catheter-related infections in critically III adults: A randomized controlled trial. JAMA-J. Am. Med. Assoc., 2009, 301, 12, p. 1231-1241.

44. Timsit J., Mimoz O., Mourvillier B., et al. Randomized controlled trial of chlorhexidine dressing and highly adhesive dressing for preventing catheter-

related infections in critically ill adults. Am. J. Respir. Crit. Care Med., 2012, 186, 12, p. 1272-1278.

45. Schwebel C., Lucet J., Vesin A., et al. Economic evaluation of chlorhexidine-impregnated sponges for preventing catheter-related infections in critically ill adults in the dressing study. Crit. Care Med., 2012, 40, 1, p. 11-17.

46. Ye X., Rupnow M., Bastide P., et al. Economic impact of use of chlorhexidine-impregnated sponge dressing for prevention of central line-associated infections in the United States. Am. J. Infect. Control, 2011, 39, 8, p. 647-654.

47. Crawford A.G., Fuhr Jr. J.P., Rao B. Cost-benefit analysis of chlorhexidine gluconate dressing in the prevention of catheter-related bloodstream infections. Infect. Control Hosp. Epidemiol., 2004, 25, 8, p. 668-674.

48. Maunoury F., Motrunich A., Palka-Santini M., et al. Cost-effectiveness analysis of a transparent antimicrobial dressing for managing central venous and arterial catheters in intensive care units. Submitted and accepted for the publication in PLoS ONE, 2015.

49. Maunoury F., Motrunich A., Ruckly S., et al. Non-homogeneous cost-effectiveness modeling of a new CHG-dressing for preventing catheter-related bloodstream infections for patients in intensive care units. International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 16th Annual European Congress, Dublin, 2013.

50. Maunoury F., Motrunich A., Fortin S. Cost-Effectiveness of the TLC-NOSF Dressing in Venous Leg Ulcers. International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 15th Annual European Congress, Berlin, 2012.

51. Maunoury F., Motrunich A., Palka-Santini M. Modeling cost-effectiveness of antimicrobial dressings for preventing catheter-related bloodstream infection: homogeneous vs non-homogeneous Markov approaches. International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 17th Annual European Congress, Amsterdam, 2014.

52. Palka-Santini M., Motrunich A., Maunoury F. Cost-effectiveness analysis of an antimicrobial transparent dressing for protecting central vascular accesses in critically ill patients versus standard transparent dressings in

France: A comparison of two modeling approaches: Decision-Tree versus Non-Homogeneous Markov Model (NHMM). International Society for Pharmacoeconomics and Outcomes Research (ISPOR), 17th Annual European Congress, Amsterdam, 2014.

53. Soufir L., Timsit J.F., Mahe C., et al. Attributable morbidity and mortality of catheter-related septicemia in critically ill patients: a matched, risk-adjusted, cohort study. Infect. Control Hosp. Epidemiol., 1999, 20, p. 396-401.

54. Renaud B., Brun-Buisson C. ICU-Bacteremia Study Group. Outcomes of primary and catheter-related bacteremia. A cohort and case-control study in critically ill patients. Am. J. Respir. Crit. Care Med., 2001, 163, p. 1584-1590.

55. O'Grady N.P., Alexander M., Burns L.A., et al. Guidelines for the prevention of intravascular catheter-related infections. Am. J. Infect. Control, 2011, 39, 4, p. 1-34.

56. Timsit J.F., Dubois Y., Minet C., et al. New materials and devices for preventing catheter-related infections. Ann. Intensive Care, 2011, 1, p. 34.

57. Eggimann P., Harbarth S., Constantin M.N., et al. Impact of a prevention strategy targeted at vascular-access care on incidence of infections acquired in intensive care. Lancet, 2000, 355, p. 1864-1868.

58. Coopersmith T.L., Zack J.E., Ward M.R., et al. Effect of an education program on decreasing catheter-related bloodstream infections in the surgical intensive care unit. Crit. Care Med., 2002, 30, p. 59-64.

59. Suetens C., Morales I., Savey A., et al. European surveillance of ICU-acquired infections (HELICS-ICU): methods and main results. J. Hosp. Infect., 2007, 65, p. 171-173.

60. Poujol I., Thiolet J.M., Bernet C., et al. Notification of healthcare-associated infections, France, 2007-2009. BEH, 2010, p. 38-39.

61. Davidoff F. Heterogeneity Is Not Always Noise. JAMA-J. Am. Med. Assoc., 2009, 302, 23, p. 2580-2586.

62. Timsit J.F., L'Hériteau F., Lepape A., et al. A multicentre analysis of catheter-related infection based on a hierarchical model. Intensive Care Med. 2012, 38, 10, p. 1662-1672.

63. Pronovost P.J., Goeschel C.A., Colantuoni E., et al. Sustaining reductions in catheter related bloodstream infections in Michigan intensive care units: observational study. BMJ, 2010, 340, p. 309.

64. Mermel L.A. What is the predominant source of intravascular catheter infections? Clin. Infect. Dis., 2011, 52, 2, p. 211-212.

65. Timsit J.F. Updating of the 12th consensus conference of the Société de Réanimation de langue française (SRLF): Catheter related infections in the intensive care unit. Ann. Fr. Anesth. Reanim., 2005, 24, 3, p. 315-322.

66. Ferreira F.L., Bota D.P., Bross A., et al. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA-J. Am. Med. Assoc., 2001, 286, 14, p. 1754-1758.

67. French National Authority for Health. Choices in Methods for Economic Evaluation: a methodological guide. HAS, 2012, 55 p.

68. Maruotti A., Rocci R. A mixed non-homogeneous hidden Markov model for categorical data, with application to alcohol consumption. Stat. Med., 2012, 31, p. 871-886.

69. Liquet B., Timsit J.F., Rondeau V. Investigating hospital heterogeneity with a multi-state frailty model: application to nosocomial pneumonia disease in intensive care units. BMC Med. Res. Methodol., 2012, 12, p. 79.

70. Hastings W.K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. Biometrika, 1970, 57, 1, p. 97-109.

71. Weinstein M.C., Fineberg H.V., Elstein A.S., et al. Clinical Decision Analysis. Philadelphia: WB Saunders Company, 1980, 351 p.

72. Beck J.R., Pauker S.G.. The Markov Process in Medical Prognosis. Med. Decis. Making., 1983, 3, 4, p. 419.

73. Sonnenberg F.A., Beck J.R. Markov models in medical decision making: a practical guide. Med. Decis. Making., 1993, 13, 4, p. 322-338.

74. Janssen J., Limnios N. Semi-markov models and applications. Norwell: Kluwer Academic Publishers, 1999, 404 p.

75. Schwebel C., Lucet J.C., Vesin A., et al. Economic evaluation of chlorhexidine-impregnated sponges for preventing catheter-related infections in critically ill adults in the Dressing Study. Crit. Care Med., 2012, 40, 1, p. 11-17.

76. Garrigues B. Étude CRRéa: Évaluation médico-économique du Coût Réel d'une journée en RÉAnimation (Medico-economic evaluation of the Real Cost of a day in ICU). [online]. Available at: http://www.snmrhp.org/Data/ModuleGestionDeContenu/application/880.pdf. (Accessed 6 May 2013).

77. Dimick J.B., Pelz R.K., Consunji R., et al. Increased resource use associated with catheter-related bloodstream infection in the surgical intensive care unit. Arch. Surg., 2001, 136, 229, p. 234.

78. Weinstein M.C., O'Brien B., Hornberger J., et al. Principles of good practice of decision analytic modeling in health care evaluation: Report of the ISPOR Task Force on Good Research Practices-Modeling Studies. Value Health, 2003, 6, p. 9-17.

79. Gelfand A.E. Gibbs sampling. J. Am. Statist. Assoc., 2000, 95, 452, p. 1300-1304.

80. Brooks S.P. MCMC Convergence Diagnosis via Multivariate Bounds on Log-Concave Densities. Technical report, University of Bristol, 1997.

81. Asmussen S., Glynn P.W. Stationarity Detection in the Initial Transient Problem. Acm. T. Model Comput. S., 1992, 2, p. 130-157.

82. Brooks S. P., Gelman A. Alternative Methods for Monitoring Convergence of Iterative Simulations. J. Comput. Graph. Stat., 1998, 7, p. 434–455.

83. Geyer, C.J. Practical Markov Chain Monte Carlo. Statistical Science, 1992, 7, p. 473-511.

84. Tierney L. Markov Chains for Exploring Posterior Distributions. Ann. Stat., 1994, 22, p. 1701–1762.

85. Gelman A., Rubin D. Inference from Iterative Simulation using Multiple Sequences, Statistical Science, 1992, 7, p. 457–511.